Institut für Biologie

Dissertation

Whole Genome Analysis of the Plant Growth-Promoting Rhizobacteria Bacillus amyloliquefaciens FZB42 with Focus on Its Secondary Metabolites

zur Erlangung des akademischen Grades doctor rerum naturalium

(Dr. rer. nat.)

Mathematisch-Naturwissenschaftliche Fakultät I

Xiaohua Chen

Dekan: Prof. Dr. Lutz-Helmut Schön

Gutachter: 1. Professor Dr. Rainer Borriss

2. PD Dr. Joachim Vater

3. Professor Dr. Jörn Piel

eingereicht: 12.2008

Datum der Promotion: 23.04.2009

Zusammenfassung

Aus der Rhizosphäre isolierte *Bacillus amyloliquefaciens* FZB42 ist ein Gram-positives sporenbildendes Bakterium. Als Biodünger besitzt er einen beeindruckenden Effekt zur Verbesserung des Pflanzenwachstums sowohl im Labor als auch in Feldexperimenten. Die Mechanismen der Pflanzenwachstumsförderung, vor allem auf molekularer Ebene, sind nach wie vor sehr schwer zu erfassen.

Um diese zu verstehen, wurde das komplette Genom von *B. amyloliquefaciens* FZB42 in dieser Arbeit sequenziert. In der Zusammenarbeit mit dem Göttingen Genomik Laboratorium, wurden verschienden e Methoden wie Shotgun, Fosmid Bibliothek, Multiplex-PCR usw. eingesetzt. Dies ist das erste sequenzierte Gram-positive Bakterium mit der Fähigkeit, Pflanzenwachstum zu fördern. Das 3916 kb große Genom, codiert mit 3693 Proteinen, ist relativ klein, dies könnte auf die fehlenden Phagen zurückzuführen sein. Die weit verbreiteten Phagen sind ein besonderes Merkmal im Genom von *B. subtilis* 168, der in enger Verwandtschaft zum FZB42 steht. Anderer Seits enthält das Genom von *B. amyloliquefaciens* FZB42 viele DNA-Inseln, in denen unikale Gene in FZB42 als Cluster gefunden wurden. Viele Gene, die möglicherweise zur Pflanzenwachstumsförderung beitragen, wurden in dieser Arbeit identifiziert. Dazu gehören Gene, die Signaltrans-duktionsweg, Interaktionen zwischen Bakterien und Pflanzen, Sekretionssystemen und Nährstofferwerb/Nutzung mit einbeziehen.

B. amyloliquefaciens FZB42 ist natürlich kompetent. Für die genetische Transformation wurde ein modifiziertes Ein-Schritt-Protokoll benutzt, das ursprünglich für *B. subtilis* 168 entwickelt wurde. Eine Reihe von Mutanten wurde über die homologe Rekombination hergestellt. Das kompetente Stadium in FZB42 kommt früher als in *B. subtilis* 168, nämlich während der späten exponentiellen Wachstumsphase. Das FZB42-Genom enthält den kompletten Satz von Genen, die für die Entwicklung der genetischen Kompetenz nötig sind. Die Mehrzahl der Kompetenz-Gene von FZB42 ist sehr ähnlich zu denen in *B. subtilis* 168. Nur das Quorum-Sensing-System von FZB42 (comQ, comX, comP) weist geringe Ähnlichkeit mit den jeweiligen Genen von *B. subtilis* 168 (36%, 31% bzw. 55%) auf.

Das FZB42 Genom birgt ein enormes Potential zur Produktion von sekundären Metaboliten. Anhand der Deletionsmutanten und der Informationen über die Genomsequenz, konnten die drei *pks* Genecluster als Bacillaene, Difficidin und Makrolactin identifiziert und ihre Biosynthesewege aufgeklärt werden. Nach über 10 Jahren Genomsequenzierung von *B. subtilis* 168 konnte zum ersten Mal das *pksX* Gen-Cluster in *B. subtilis* 168 Bacillaene zugeordnet werden. Genetische Manipulationen wurden durchgeführt, um die Funktionen der *trans*-AT Domänen und der Modifikationsdomänen in den PKS-Gen-Clustern zu erklären.

Zusätzlich zu diesen drei PKS-Gen-Clustern wurden sechs riesige Gen-Cluster identifiziert, die für die Produktion von nicht-ribosomalen synthetisierten Peptiden zuständig sind,. Mit Ausnahme von vier Gen-Clustern in *B. subtilis* 168 (Surfactin, Fengycin, Bacillibactin und Bacilysin), sind Bacillomycin D und ein hypothetisches Tripeptid einzigartig im Genom der FZB42. FZB42 kann kein bekanntes ribosomal synthetisiertes Bacteriocin produzieren kann. Gleichzeitig beinhaltet sein Genom ein Gen-Cluster, das wahrscheinlich für die Produktion eines neuartigen Bacteriocins verantwortlich ist. Insgesamt nutzt FZB42 mehr als 8,5% seiner gesamten Genomkapazität, um antimikrobielle Stoffe zu synthetisieren. Die eindrucksvolle genetische Kapazität zur Herstellung von antagonistischen sekundären Metaboliten ermöglicht es FZB42, nicht nur erfolgreich neben konkurrierenden Organismen innerhalb seiner natürlichen Umgebung zu überleben, sondern auch Pflanzen gegen pathogene Bakterien und Pilze zu schützen.

Abstract

Bacillus amyloliquefaciens strain FZB42, a Gram-positive endospore forming bacteria, was isolated from the rhizosphere. As a biofertilizer, it has been already shown an impressed the effect in the laboratory as well as in field experiments to enhance plant growth. However, the mechanisms of plant growth promotion, especially at the molecular biological level, remain elusive. To comprehensively understand these mechanisms, the complete genome of *B. amyloliquefaciens* FZB42 was sequenced in this work by using variable tools like shotgun, fosmid library, multiplex PCR etc. in co-operation with Göttingen Genomic Laboratory. This is the first sequenced representative of Gram-positive bacteria with biocontrol and plant growth-promoting activity. The genome of 3,916 kb containing 3,693 protein-coding sequences is relatively small; this could be explained by the absence of extended phage insertions which are typical for the closely related *B. subtilis* 168 genome but contains. On the other hand, several DNA islands where unique genes in FZB42 were found clustered. Many candidate genes that may contribute to the plant growth promotion were identified in this works. These include genes that involved in signal transduction pathways, plant-bacterial interactions, secretion systems and nutrient acquisition/utilisation.

B. amyloliquefaciens FZB42 is naturally competent and amenable for genetic transformation using a modified one-step protocol, originally developed for *B. subtilis* 168. This property was used to generate a set of mutants impaired in many distinct functions by a gene replacement strategy *via* homologous recombination. FZB42 exhibited its maximal competence earlier than *B. subtilis,* during late exponential growth. Not surprisingly, the FZB42 genome harbors the complete set of genes necessary for development of genetic competence. The majority of competence genes is highly homologous to their counterparts in *B. subtilis* 168, but the genes that control the competence quorum-sensing system of FZB42 (*comQ, comX, comP*) exhibit low similarity to the respective genes of *B. subtilis* 168 (36%, 31%, and 55% respectively).

The FZB42 genome harbors an enormous potential to produce secondary metabolites, which was investigated in the second part of this work. According to the information of the genome sequence and the mutants with disruption in these three pks gene clusters, they could be identified as bacillaene, difficidin and macrolactin and their biosynthesis pathways could be elucidated. For the first time, bacillaene could be assigned as the product for pksX gene cluster in B. subtilis 168, after over 10 years completely sequencing the genome of BS168. Genetic manipulation was carried out to investigate the trans-AT domains and some modification domains in the pks gene clusters. In addition to these three *pks* gene clusters, six

giant gene clusters directing synthesis of pe ptides were identified. Except for four gene clusters present in BS168, b acillomycin D and a hypo thetical tripeptide are unique in the genome of FZB42. A remarkable feature of the FZB42 genome is that it does not produce any known ribosomally synthesized bacteriocin, whereas a gene cluster probably responsible for production of a new bacteriocin was identified in this work. In total, FZB42 utilizes more than 8.5% of its whole genom ic capacity to synthe size antimicrobial substances. The impressive genetic capacity to produce antagonistic acting secondary metabolites not only enables FZB42 to cope successfully with com peting or ganisms within its natura l env ironment, but also to protect plants from pathogenic bacteria and fungi.

Schlagwörter:

Bacillus amyloliquefaciens FZB42, Genome, Polyketide, Lipopeptide, Antibiotikum

Keywords:

Bacillus amyloliquefaciens FZB42, genome, polyketide, lipopeptide, antibiotic

Table of contents

Acknowledgements

This work could not be completed without the support, co-operation and guidance of many people.

I would like to thank Professor Dr. Rainer Borriss, my supervisor who provided me the opportunity to work with these interesting topics, the brilliant supervision and top laboratory equipments. His wisdom and knowledge inspired and motivated me during the whole work.

I thank also all members of our group, for their help and for making the work in the laboratory a pleasure, especially to Alexandra Koumoutsi (ex-member of our group) for many useful discussions and Romy Scholz who established the HPTLC and bioautography methods in our laboratory.

Sincere thanks are given to all members in G2l Göttingen. Special thanks to: Professor Dr. Gottschalk (G2L) who gave us the possibility to sequence the whole genome of FZB42, which is one of the most important parts of my dissertation; Axel Strittmatter (G2L Göttingen) and Anke Henne (G2L Göttingen) who coordinated the sequencing work in G2L; and Iwona Decker who did the most PCR sequencing work in the genome project.

Thanks to Dr. Heiko Liesegang (G2L Göttingen) and Markus Loose (computer science, Humboldt University of Berlin) for their bioinformatics support that accelerate date processing. Dr. Rainer Cramm and Dr. Anne Pohlmann (AG Microbiology, Humboldt university of Berlin) provided their annotation program Genesoap that facilitated the annotation work and usefully discussed about the general problems of genome projects.

I would sincerely thank to Dr. Joachim Vater who was always engaged in identifying different secondary metabolites produced in FZB42. His long-time experience in the analytical chemistry was worthwhile for analysing the secondary metabolites of FZB42. Thanks again for his willingness to be a referee for my thesis.

I wish to express my gratitude to Professor Dr. Jörn Piel from Rheinische Friedrich-Wilhelms-Universität Bonn for beeing my referee and for many valuable discussions about analyzing polyketides. Special thanks to his co-worker Jana Moldenhauer who is open for the collaboration and experimentally investigated the biosynthesis pathway of bacillaene.

I would like to give special thanks to professor Dr. Roderich Süßmuth and his co-worker Kathrin Schneider for their fruitful collaboration in the project to identify the polyketides produced in FZB42 using LC ESI-MS and to characterize the starter and extender units of macrolactin using NMR MS.

A big thank to Lilia Costa Carvalhais, Eva Garcia Gonzalez, Kinga Kierul, Aruljothi Mariappan and Paul Dennis who carefully corrected my English writing.

Finally, I want to thank my family and all my friends for always being there.

Abbreviation

A	adenylation domain
aa	amino acid
ACP/T	acyl-carry protein
Ala	alanine
Ap	ampicillin
ArCP	aryl carrier protein
AT	acyl-transferases
ATP	adenosine-5'-triphosphate
bac	bacilysin
bae	bacillaene
BAM	*Bacillus amyloliquefaciens*
BCIP	5-Bromo-4-chloro-3-indoxylphosphate
bmy	bacillomycin
BS	*Bacillus subtilis*
BSA	Bovine Serum Albumin
C	condensation domain
CD	conserved domain
cm	chloramphenicol
Cy	cyclization
DNA	deoxyribonucleic acid
dfn	difficidin
DH	dehydratase
DHB	2,3-dihydroxybenzoate
DIG	digoxigenin
EH	enoyl-CoA hydratase
em	erythromycin
EPSs	exopolysaccharides
ER	enoyl reductase
FAS	fatty acid
fen	fengycin
fig	figure
Gly	Glycin
h	hours

HMGS	3-hydroxy-3-methylglutaryl-CoA synthase
HPLC	high performance liquid chromatography
HPLCESI-MS	high performance liquid chromatography coupled electrospray ionisation mass spectrometry
IAA	indole-3-acetic acid
IPTG	isopropyl-ß-D-thiogalactopyranoside
ISR	induced systemic resistance
kb	kilo basepair
kDa	kilo Dalton
km	kanamycin
KR	ketoreductase
KS	β-ketoacyl synthase
LB	Luria-Bertani
Lys	lysine
MALDI-TOF-MS	matrix assisted laser desorption ionization time of flight mass spectrometry
min	minutes
mln	macrolactin
MT/M	methylase
MUMs	maximal, unique, matching sequences
NMR-MS	nuclear magnetic resonance mass spectrometry
NRPS	nonribosomal peptide synthetase
OD	optical density
ORF	open reading frame
PCP	peptide carrier protein
PCR	polymerase chain reaction
PGPR	plant growth-promoting rhizobacteria
PKS	polyketide synthase
Ppant	4′-phosphopantetheinyl
PPTase	4’-phosphopantetheinyl transferase
RNA	ribonucleic acid
rpm	rotations per minute
RT-PCR	reverse transcription polymerase chain reaction
SAM	S-adenosyl-methionine

Ser	serine
SMR	small multidrug resistance
srf	surfactin
TAE	Tris / acetic acid / EDTA
TBE	Tris/Boric acid/EDTA
TE	thioesterase
Thr	threonine
TMH	transmembrane helix
Trp	tryptophan
Tyr	tyrosine
UPP	undecaprenyl pyrophosphate
VOCs	volatile organic compounds
X-Gal	5-Bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

1 Introduction

Microbial genome sequencing and analysis is a rapidly expanding and increasingly important trend in microbiology. Before whole genome analysis became available, life science based on a reductionism principle, dissecting cells and systems into their fundamental components for further study. Recent studies on whole genomes and whole genome comparisons in particular give us a possibility to receive more information leading to better understanding of the general feature of the life.

The explosion of the world population and the demand of high quality of life require a higher productivity of agriculture leading to the enhanced use of chemical fertilizers and pesticides. Nowadays alternative possibilities are explored to overcome the problems of environmental pollution and nutritional deficiency caused by a rapidly growing world population. Utilization of environmental microorganisms, which can support plant growth and suppress growth of variable plant pathogens, seems to be one of the most feasible approaches. Recently, different *Bacilli*, which show benefit for plant growth, were isolated worldwide from nature. *Bacillus amyloliquefaciens* FZB42 (abbreviated as FZB42 in this work), isolated by ABITEP Berlin (former FZB Berlin), belongs to this category. Bioformulations containing spores makes *Bacilli* more convenient for storage and transportation than Gram-negative bacteria like *Pseudomonas*. Unfortunately, knowledge about the mechanisms of plant growth promotion induced by Gram-positive bacteria is rather scarce compared to their Gram-negative counterparts.

Genomic sequencing *Bacillus amylolique faciens* FZB42 allows us now to understand comprehensively the mechanisms of plant growth promotion exerted by Gram-positive rhizobacteria on the molecular level.

1.1 The Genus of Bacillus

Bacillus is a Gram-positive and rod-shaped bacterium with low GC-content, which is ubiquitous and broadly adapted to grow in diverse environments, either terrestrial or aquatic. *Bacillus* can form highly resistant dormant endospores in response to nutrient deprivation or other environmental stress conditions [1]. Spores can remain viable for extended periods of time. Under favourable conditions, the spore starts to germinate and converts to a vegetative cell again [2].

The genus *Bacillus* can be divided in at least five different groups based on 16S rRNA sequence analysis [3, 4]. *Bacillus subtilis* 168 (abbreviated as BS168 in this work), obtained from *Bacillus subtilis* Marburg strain after X-ray irradiation [5], is one of the best-understood prokaryotes in terms of molecular biology and cell biology [6]. Its whole genome sequence

[7] and superb genetic a menability provide the powerf ul tools requ ired to inve stigate a bacterium from di fferent possib le aspects. The or ganism is a m odel for inv estigating differentiation, gene/protein regulation and ce ll cycle ev ents in Gram -positive bacteria. However, random mutations introduced in past in BS168 by X-ray irradiation led to silencing of numerous genes that are activ e in environmental strains. For this re ason, it is nece ssary to investigate its closes t relatives for understanding the role of these genes and the relationship between *Bacillus* and its environm ent. Extensive work has been already ca rried out in these areas, but one of the m ost dif ficult challen ges is that the majority of the investig ated environmental strains are highly recalcitrant to transform ation hindering applying of genetic approaches.

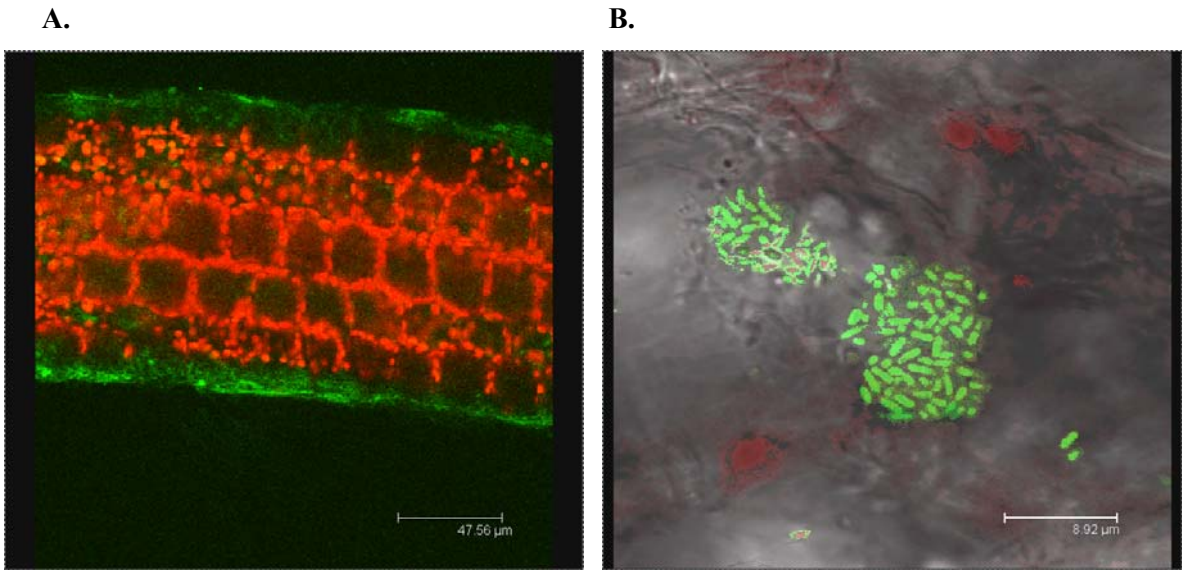


Figure 1. Gfp labelled Bacillus amyloliquefaciens FZB42 colonizes on a root of Lemna minor: A. The whole r oot (r ed) is cover ed by a shield of B. am yloliquefaciens FZ B42 (gr een). B. T wo colonies of B. amyloliquefaciens F ZB42 migra ted into the inter -cellular space in the leaves. (Kindly provided by Ben Fan, Bacterial Genetics, Humboldt University of Berlin)

B. amyloliquefaciens, discovered first tim e by Fukum oto (1943), was named according to its capability to produce alpha am ylase. The debate abou t its nom enclature had con tinued for some decades, after *B. amyloliquefaciens* was separated from the species *B. subtilis* in 1988 due to its dif ferent 16s ribosom al RNA seque nce [4]. *Bacillus amyloliquefaciens* FZB42 (**Figure 1**), isolated from soil, dif fers from the model strain BS168, by its ability to enhance plant growth and to produce various secondary metabolites with antibio tic activity [8, 9]. Until to th e last decade, m olecular and cell biological knowledge about m echanisms underlying plant growth prom otion and biosynthe sis of these secondary m etabolites is poorly understood.

1.2 Plant growth-promoting rhizobacteria

Plant growth-promoting rhizobacteria (PGPR) were first defined by Kloepper and Schroth [10] to describe soil bacteria that colonize the roots of plants and exert beneficial effects on plant growth by acting as bio-fertilizers, bio-pesticides, phyto-stimulators and bio-remediators.

In general, colonization of rhizobacteria is the most essential and primary step in the process of plant growth promotion. Rhizobacteria live on plant roots or reside in the rhizosphere, a soil zone spanning a few millimeters around roots, where they fed by root exudates. The following steps are implicit in colonization process: (1) proliferating in the spermosphere (region surrounding the seed), (2) attaching the root surface, and (3) colonizing the developing root system in presence of seed/root exudates [11]. However, a variety of bacterial features like quorum sensing, motility, chemotaxis, production of specific cell surface components and ability to use specific components of root exudates, which could contribute to this process, were analyzed and a few genes have been identified. Most of this work was focused on Gram-negative strains belonging to *Pseudomonas spp.* [12, 13].

PGPR can directly enhance plant growth by numerous mechanisms: (1) Fixation of atmospheric nitrogen that is transferred to the plant [14], (2) production of siderophores that chelate iron and make it available to the plant root, (3) solubilization of minerals such as phosphorus [15, 16], (4) Production of phytohormones by rhizobacteria [15, 17, 18]. (5) Secretion of antibiotics into the rhizosphere can protect plants directly from pathogenic microorganisms or competitively suppress them by colonization [9, 19, 20].

Since Kloepper and Schroth [10] have introduced the concept of PGPR in 1978, the most investigated subjects for PGPR have been Gram-negative strains like *Pseudomonas* [13]. Also at present, many mechanisms of PGPR in Gram-positive and spore forming bacteria, like *Bacillus*, remain poorly understood.

1.3 Secondary metabolites

Secondary metabolites are those chemical compounds in organisms that are not directly involved in normal growth, development or reproduction of organisms. They are often species-specific or found in only a small set of species in a narrow phylogenetic group. Many important secondary metabolites produced by PGPR are involved in plant growth promotion [9, 13, 21-24]. One of the most interesting aspects in the mechanisms of plant growth promotion is bio-control. Many secondary metabolites show activities against pathogenic microorganisms and parasites. Discovering new antibiotics and unraveling the mechanisms of antibiotic resistance are interesting research fields, not only because of their utilization in bio-

control practices [24, 25] but also in keeping down increased opportunities for development of infections and multi-resistance against known antibiotics in clinics [26, 27]. The potential of *Bacilli* to produce antibiotics is known for more than 50 years. In average, 4-5% of the *Bacillus subtilis* genome contributes to the production of antibiotics [28]. Antibiotics are divided into two groups depending on their biosynthesis pathway: 1) ribosomally synthesized peptides that are normally modified post-translationally, like bacteriocins; 2) nonribosomal synthesis via mega-enzyme complexes, like polyketides and lipopeptides.

1.3.1 Ribosomally synthesized antibiotics: bacteriocins

Bacteriocins are ribosomally synthesized peptide antibiotics. Both, Gram-negative and Gram-positive bacteria can produce and export bacteriocins in to the environment in order to antagonize their competitors. Bacteriocins are divided into two classes depending on whether they possess modified amino acids. Bacteriocins with modified amino acids belong to class I and the ones without modification belong to class II [29-31].

1.3.1.1 Modified bacteriocins (class I)

Bacteriocins with modified amino acids are designated lantibiotics because of the thioether amino acids lanthionine and methyllanthionine [32]. The pre-peptide contains an N-terminal leader sequence, which seems to be required for modification of the C-terminal structural region and is removed by proteolysis to produce the mature lantibiotic. Lantibiotics were divided into two subgroups based on their structure [33]: Type A are elongated, screw-shaped and amphiphic; Type B are more globular in structure. Type A lantibiotics interact with the membrane of susceptible cells and form transient voltage-dependent pores [34], whereas some of type B lantibiotics inhibit functions of some enzymes.

Lantibiotics were already found in many *Bacilli* (**Figure 2**). As a type A lantibiotic, subtilin (32 aa) was first identified in *B. subtilis* ATCC6633. Its structure is similar to nisin [35, 36]. Ericin A (32 aa) and ericin S (29 aa), identified from *B. subtilis* A1/3, belong to type A. The structural genes for ericin A/S are similar to subtilin, but ericin A shows a higher antibiotic activity than ericin S [37]. The gene cluster for biosynthesis of sublancin (type A lantibiotic with 37 aa) is in the region of prophage sp β2 [38]. Mersacidin inhibiting biosynthesis of peptidoglycan and transglycosylation [39-41] is a representative of type B lantibiotics. Based on the structural analysis, subtilosin (35 aa) is the only known circular lantibiotic in *Bacilli* [42]. The three post-transcriptional thioether linkages between Cys and Phe/Thr lead to classify subtilosin to the class of lantibiotics. However, oxidative linkage between the sulphur of the cysteine and the α-carbon of another amino acid residue is a novel type of post-transcriptional modification, allowing to classify subtilosin in an own group I [42].

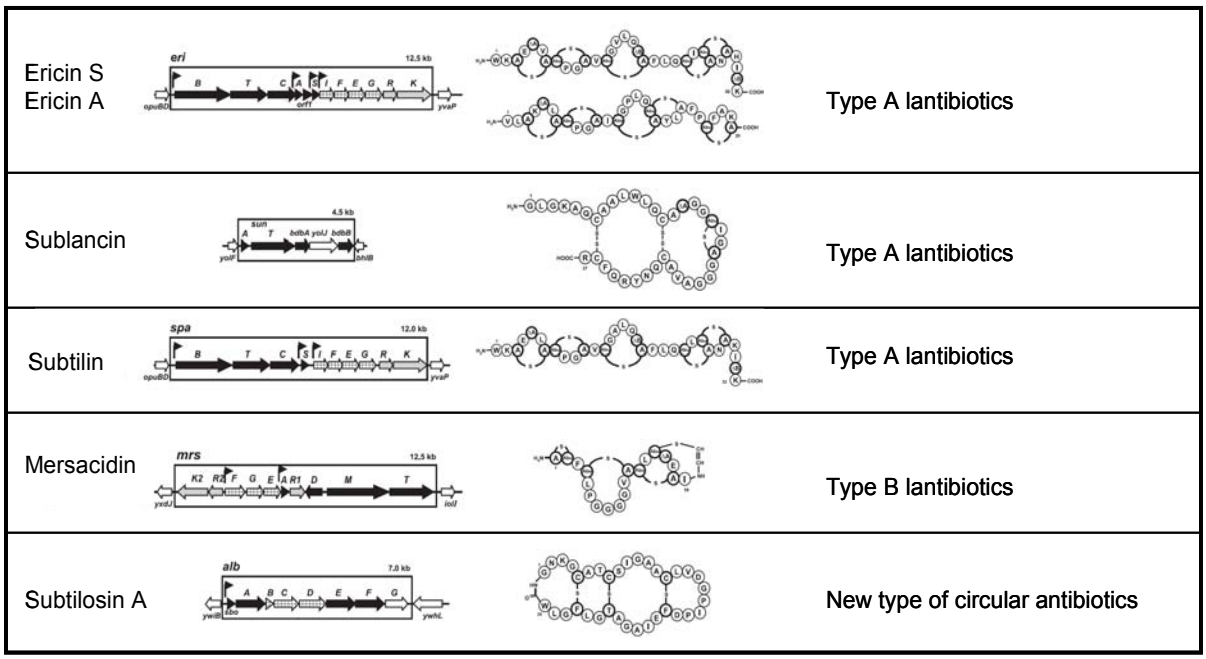


Figure 2. Summary of lantibiotics produced by Bacilli (modified after T. Stein [28]).

1.3.1.2 Unmodified bacteriocins (class II)

Unmodified bacteriocins do not contain modified amino acids, are cationic and hydrophobic peptides with a length of 20 to 60 amino acids. Class II bacteriocins are mainly active against low GC-content Gram-positive bacteria. The inhibition spectrum is rather narrow, limited to species or strains related to its producer [43].

Bacteriocins are divided into three subclasses, IIa, IIb and IIc, according to their primary structures. Among the subclass IIa (antilisterial pediocin-like bacteriocins), many of the m share high similarities in their primary structure. Particularly, they have a highly conserved N-terminal part including the YGNGV consensus motif [44]. Bacteriocins of the subclass IIb (two-peptide bacteriocins) contain heterodimer peptides that show no significant similarity in their primary structure. The single peptide has normally no activity. The subclass IIc represents leaderless peptide-bacteriocins and the subclass IId circular bacteriocins [45].

Class II bacteriocin has not been identified in any *Bacillus* strain so far.

1.3.2 Non-ribosomally synthesized antibiotics: Lipopeptides and polyketides

Non-ribosomally synthesized peptides and polyketides show a broad range of structural diversity and bioactivity. They include many important antibiotics [46-48], antitumor agents [49-52], immunosuppressive agents [53] and siderophores [54, 55]. In spite of their structural heterogeneity, polyketides and non ribosomal peptides represent two families of natural products biosynthesized by similar multi-modular enzymes acting in assembly line arrays.

PK-NRP hybrids are compounds biosynthesized by mega-enzyme complexes that contain both PKS and NRPS modules.

1.3.2.1 Nonribosomal lipopeptides

1.3.2.1.1 Mega enzyme complex of NRPS

The nonribosomal peptide antibiotics share a common mode of synthesis, the multicarrier thiotemplate mechanism. The assembly of the multifunctional proteins of peptide synthetases involved in nonribosomal peptide biosynthesis is reflected by its genetic organization following the co-linearity rule. This mechanism resembles the amino-acylation of tRNA synthetases during ribosomal peptide biosynthesis. Each elongation cycle in non-ribosomal peptide biosynthesis needs at least three core domains: adenylation domain (A), peptidyl carrier protein (PCP) domain (also known as thiolation (T) domain) and condensation (C) domain (**Figure 3**). Adenylation domains select appropriate amino acid monomers from the cellular pool, activate it with ATP to aminoacyl-AMPs and transfer it to the phospho pantetheinyl group of PCP domain. The adenylated amino acid substrate is connected to the PCP domain in thioester bound. Condensation domains (C) catalyze the formation of a new peptide bond between the aminoacyl thioester attached to the PCP domain of the same module and that of the preceding module. The C domain can be substituted by a Cy domain that catalyses both condensation and the intra-molecular hetero-cyclization of serine, cysteine or threonine [56, 57] (**Figure 3**). The initiation module of NRPS normally lacks a C domain and the last module usually contains an extra thioesterase (TE) domain (or an alternative domain, such as a reductase) to release the fully assembled peptide from the PCP domain of the final module (**Figure 3**). The TE domain cleaves the covalently tethered peptide from the NRPS by catalyzing the nucleophilic attack of water (resulting in a linear peptide) or an internal nucleophile (resulting in a cyclic peptide).

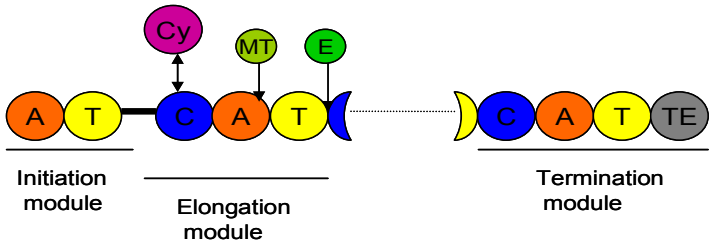


Figure 3. Module organization of NRPS: the initiation module contains normally only domain A and T. The number of elongation modules is dependent on the number of amino acids assembled in the lipopeptide. The TE domain exists at the end of the termination module. The C domain can be substituted by a Cy domain. The epimerisation domain (E) and the methyltransferase domain (MT) are involved in the modification of amino acids.

Except these essential domains, some extra enzymes exist in some modules. An epimerization (E) domain, following a PCP domain and located at the C-terminus of a module, catalyses racemization or epimerization (L- to D-configured amino acids), either before or after peptide bond formation [58, 59] (**Figure 3**). Methylation of the nitrogen of the amino group is achieved by a methyltransferase (MT) domain, which is inserted into the A domain (**Figure 3**). NRPS modules can function as monomers or dimers. Dimeric structure is important for synthesis to proceed *in trans* as well as *in cis* [60, 61].

1.3.2.1.2 Features of nonribosomal lipopeptides and their representatives in Bacilli

Cyclic lipopeptides such as surfactin, fengycin, and iturin-like antibiotics are widely produced by *Bacillus subtilis* and related strains (**Figure 4**).

Surfactin is a cyclic heptapeptide with an LLDLLDL chiral sequence linked to a β-hydroxy fatty acid with 13 to 15 carbon atoms via a lactone bond. Surfactins exist widely in *Bacilli*. The *srfA* operon is required for the nonribosomal biosynthesis surfactin. Itis composed of the four genes, *srf*AA (three modules), *srf*AB (three modules), *srf*AC (one module), and *srf*AD (**Figure 4**). Surfactin exerts its antimicrobial and antiviral effect by altering membrane integrity [62].

Fengycin and the closely related plipastatin are cyclic lipodecapeptides containing a β-hydroxy fatty acid with a side chain length of 16 to 19 carbon atoms. Four D-amino acids and ornithine (a nonproteinogenic residue) have been identified in the peptide portion of fengycin (**Figure 4**). The operon *fen*ABCDE is responsible for the biosynthesis of fengycin. FenA, FenB and FenC encode two modules, FenD encodes three modules and FenE encodes only one module (**Figure 4**). Fengycin is specifically active against filamentous fungi and inhibits phospholipase A [63].

The members of the iturin family are amphiphilic compounds that are composed of seven amino acid residues including an invariable D-Tyr2, with the constant chiral sequence LDDLLDL connected to a C14-C17 aliphatic beta-amino acid (Figure 4). The iturins, the mycosubtilin and the bacillomycins belong to this family. The gene organization of the iturin family is very similar: malonyl-CoA transacylase is the first gene in the gene cluster, followed by three genes containing 1, 4 and 2 modules for the assembly of the growing peptide chain of amino acids respectively (Figure 4) [8, 64-66]. They exhibit strong hemolytic and antifungal activities against a wide variety of fungi, but their antibacterial activities are restricted. The antifungal activity is related to the interaction of the lipopeptides with the cytoplasma membrane and change of its permeability [8].

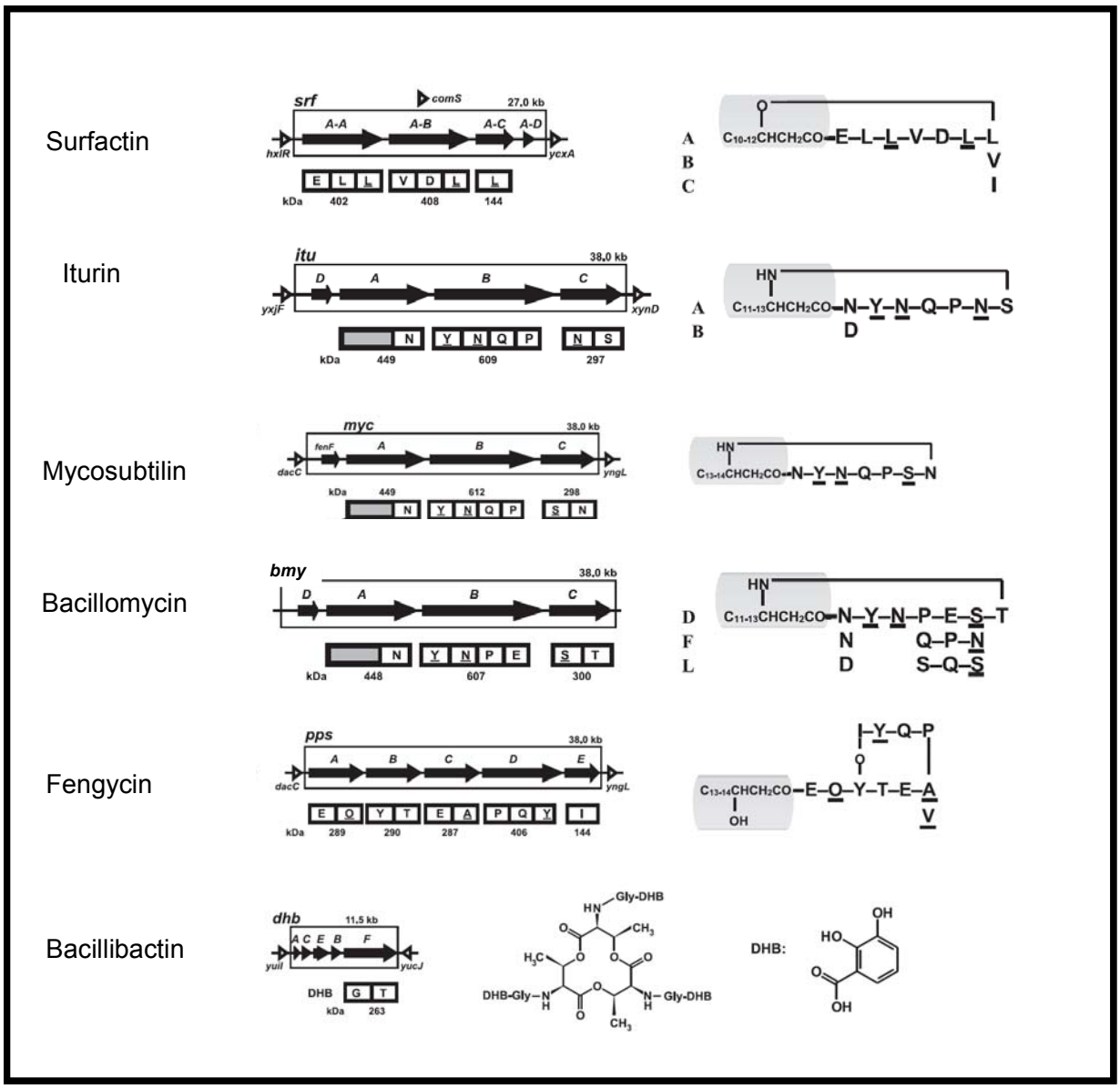


Figure 4. Summary of sfp-dependent nonribosomally synthesized peptides in Bacilli. D-amino acids are underlined (modified after T. Stein [28]).

Bacitracin is a cyclic dodecalpep tide antib iotic. It can be found in *B. lichen iformis* and *B. subtilis* [67]. BceA contains f ive modules, with an intern al epimerization domain attached to the fourth; BceB contains tw o modules and a carboxy-term inal epim erization dom ain; BceC contains five m odules with addi tional internal epim erization domains attached to the second and fourth. A carboxy-term inal putative thioeste rase dom ain was also detected in BceC. A putative cyclization dom ain wa s found in BceA that m ay be involved in thiazoline ring formation [68]. Bacitracin binds tightly to und ecaprenyl pyrophosphate (UPP) and prevents the dephosphorylation of UPP nece ssary for the recycling of the lipid carrier , thereby ultimately inhibiting cell wall polymer biosynthesis [69].

Iron is a co factor of essential cellu lar pro cesses in nearly all m icroorganisms. Because the bioavailability of iron is extrem ely lim ited in most of the biospheres, a great number of bacteria including im portant pathogens produ ce ferric iron-chelati ng compounds known as siderophores to gain access to various iron sources. Three m ain siderophore groups can be classified: catecholate, hydroxamate and carboxylate siderophores. The catecholic siderophore bacillibactin is produced by dif ferent *Bacillus* strains such as *B. subtilis* , *B. anthracis* , *B. cereus* and *B. thuringiensis*. Gene cluster *dhbACEBF* encodes genes for the biosynthesis of bacillibactin. Its precursor 2,3-dihy droxybenzoate (DHB) was synthesized by *dhbAC* and ligase with glycin and threonine to for m a tri-peptide intermediate by utilizing the tri-modular NRPS system DhbEF (**Figure 5**). DhbB is responsible for circulating three tri-peptides to form the bacillibactin (**Figure 5**) [55].

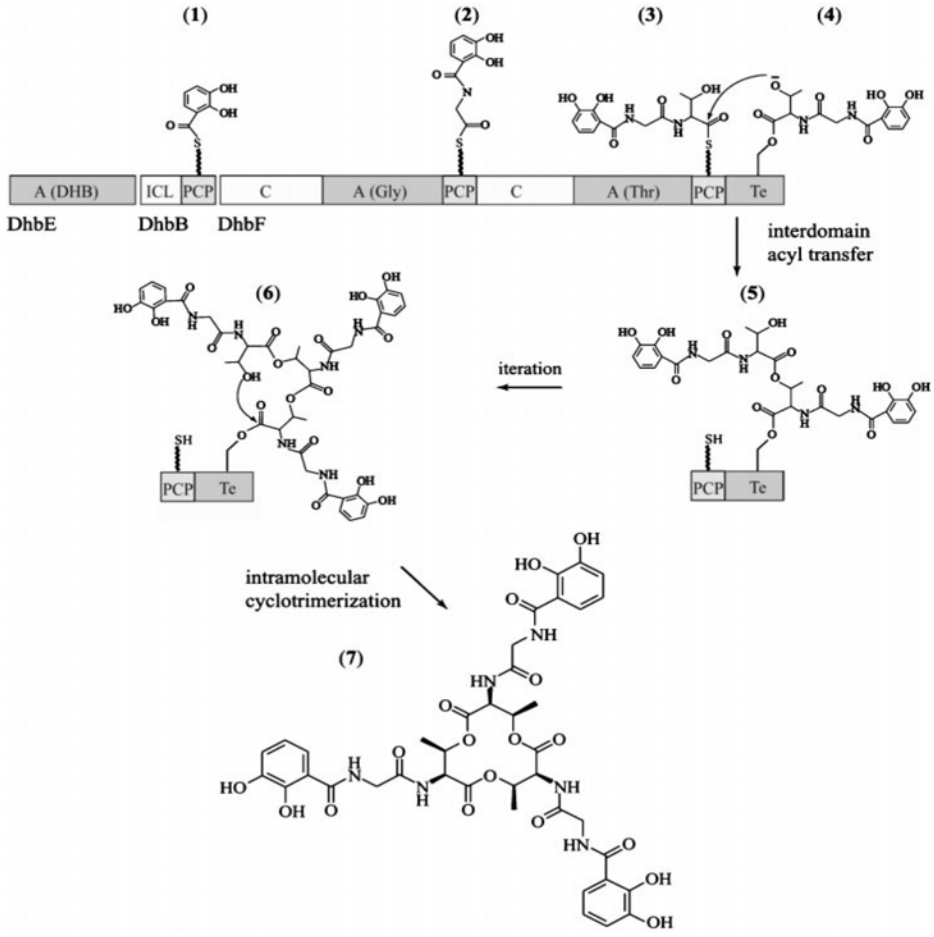


Figure 5. Model for the assembly of bacillibactin: (amino) acids are activated and are bound as acyl-S-Ppant interme diates on the corr esponding carrier pr oteins. Condensation r esults in formation of the DHB- Gly-Thr unit on the la st PCP domain. The peptide chain is then transferred to a conser ved serin e residue of t he ter minal Te domain. Te catalyzes th e condensation of thr ee DHBGly-Thr units and fi nally r eleases bacillibactin by intramolecular cyclization. ICL is isochorismate lyase [55].

The detailed knowledge about the biosynthesis of surfactin [70], fengycin [63] and iturins [8, 66] allows bioengineering of novel lipopeptides.

1.3.2.2 Polyketides

1.3.2.2.1 Mega enzyme complex of PKS

PKSs can be classified into three types based on their protein architectures. Type I PKSs are multi-domain proteins that modularly or iteratively prolong the polyketide chain by using an acyl carrier protein (ACP) [71], while type II PKSs are comprised of a set of iteratively used enzymes, each expressed from a distinct gene [72]. The aromatic polyketides are synthesized by either the iterative type I or type II. Type III PKSs are homodimeric proteins that are similar to the large family of plant chalcone synthase-like proteins. Here the chain growth is not dependent on the phosphopantetheinyl group of ACP [73].

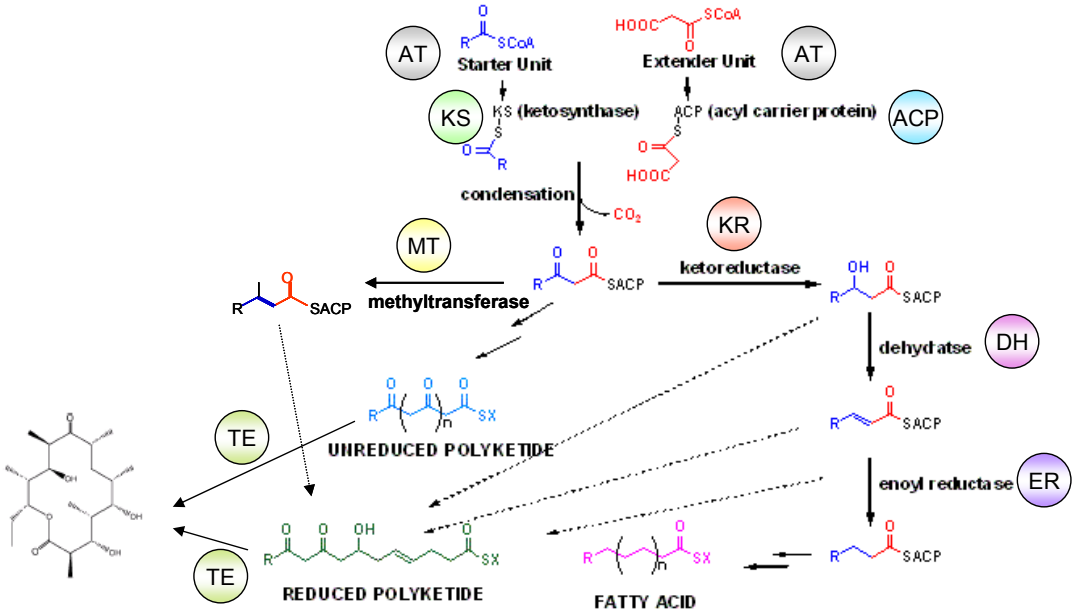


Figure 6. Roles of various domains in the biosynthesis of polyketides in comparison with fatty acid biosynthesis [76].

Similar to NRPS, modular type I PKSs follow an assembly-line mechanism. Every module is responsible for one round of condensation and reduction. The "mini module" consists of acyl transferase (AT) domain, acyl carrier protein (ACP) domain and β-acyl-ketosynthase (KS) domain. The majority of the AT domains display tight substrate selectivity. They can load specifically substrates like malonyl-CoA, methlymalonyl-CoA, ethylmalonyl-CoA, and so on. These substrates direct biosynthesis of polyketides in a stepwise manner from simple 2-, 3-, and 4-carbon building blocks [74, 75]. The AT domain transfers the substrate to the

phosphopantetheinyl group of ACP and the KS domain catalyzes a decarboxylative Claisen condensation to connect the extender unit to the growing polyketide chain. Finally, a thioesterase (TE) domain releases the polyketide from the last ACP domain and catalyses cyclization of the polyketide (**Figure 6**) [76].

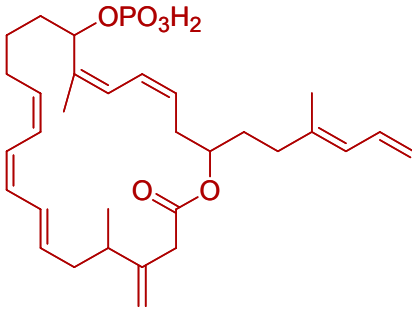
Unlike saturated fatty acid biosynthesis, PKSs display different level of keto to thioester reduction by selectively utilizing β-keto-acyl reductase (KR) domain, hydroxyacyl dehydratase (DH) domain and enoyl reductase (ER) domain (**Figure 6**). KR domain carries out the first reduction to an alcohol. DH domain can further eliminate a water to form a double bond. ER domain catalyses the final reduction to form fully saturated single bond.

Additionally, PKSs can perform a methylation on carbon and oxygen with S-adenosyl-methionine (SAM)-dependent methyltransferases (MT). Carbon-specific methyltransferases couple the methyl group onto the activated α-carbon of the β-ketothioester intermediate, either during or after assembling the natural product (**Figure 6**) [77].

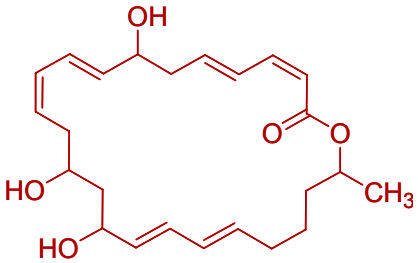
trans-AT PKSs are a special architecture for type I PKSs, where AT domains are not present in every module. Iteratively operating discrete AT(s) load the extender units to AT-less type I PKS(s). Such *trans*-ATs have been already identified in leinamycin, pederin and chivosazol biosynthetic assemblies [78-80].

1.3.2.2.2 Features of polyketides and their representatives in Bacilli

Polyketides exist widely in living cells, from bacteria to mammalians, but can exert different functions in living processes. Many polyketides are well-known compounds as antibiotics such as erythromycin A (a broad spectral antibiotic). In addition to the field of antimicrobials, further compounds have been approved as immunosuppressants (rapamycin, ascomycin) [81, 82], as anticancer compounds (bleomycin, dactinomycin) [83, 84] and as antihelmintic agents (milbemycin) [85].



Difficidin



Macrolactin

Figure 7. the structures of difficidin and macrolactin

All these polyketides are produced by *Streptomyces*, a representative for polyketide producer. Meanwhile there are only three known polyketides (bacillaene [86], difficidin [87, 88] and macrolactin [89, 90]) produced by *Bacilli*. From these three polyketides, only the structures of difficidin [87] and macrolactin [90] have been determined (**Figure 7**). No information about their genetic organization of PKS and biosynthesis pathway was available. The gene cluster *pks*X in *Bacillus subtilis* 168 [7] has been thought to be responsible for the production of difficidin, however no polyketide could be detected in *B. subtilis* 168. The reason why BS168 does not produce any polyketides is not explained yet, but one of the most possible hypothesis is that some important genes involved in the biosynthesis are dysfunctional because of X-ray irradiation [91, 92].

1.3.2.3 4'-phosphopantetheinyl transferase (PPTase)

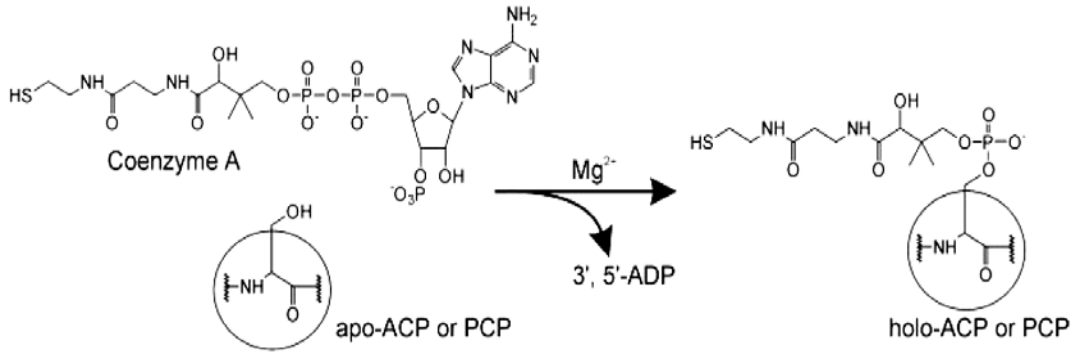


Figure 8. Activity of PPTases to convert apo- ACP/PCP to holo-ACp/PCP by transferring the 4-PPant moity of CoA on the conserved serine residue of apo-ACP/PCP [94]

Nonribosomal peptide synthetases, polyketide synthases and fatty acid synthases require posttranslational modification of their constituent peptidyl carrier protein (PCP) domains or acyl carrier protein (ACP) domains to become catalytically active. The inactive *apo*-proteins are converted to active *holo*-proteins by phosphopantetheinyl (Ppant) transferases that can transfer the 4′-phosphopantetheinyl moiety of coenzyme A to the side chain hydroxyl of a conserved serine residue in each apo-protein domain (Figure 8) [93]. The newly introduced -SH of the Ppant prosthetic group acts as a nucleophile for acyl-CoA or its derivates for fatty acids and polyketides, or aminoacyl-AMPs for peptide biosynthesis [94].

The first Ppant transferase being cloned and characterized was AcpS from *Escherichia coli*, responsible for *apo* to *holo* conversion of fatty acid synthase. Not AcpS, but EntD (another phosphopantetheinyl transferase in *E. coli*) can modify the inactive *apo*-PCPs in EntF to the *holo*-PCPs, which indicates that AcpS has substrate specificity. By searching in the database and by functional analysis, Walsh´s group could identify some new phospho pantetheinyl

transferases, which include Sfp (*Bacillus subtilis*), EntD (*E. coli*) and Gsp (*Bacillus brevis*) [95]. Surprisingly, initial searches of sequence databases did not reveal any protein with significant peptide sequence similarity with AcpS [95].

Phosphopantetheinyl transferases are grouped into three families. The first group is the AcpS type of *E. coli*. PPTases of this type comprise about 120 aa. They are found in almost all microorganisms and are involved in the modification of fatty acid ACP. They were shown to activate ACP of PKS type II [96]. Sfp like PPTases, the second group, are 240 aa length and possess an extremely wide substrate specificity, like PCPs of NRPS and ACPs of FAS and PKS [95]. The third group is formed by integrated PPTases that can be found as the C-terminal domain of the multifunctional FAS2 [97].

The choice of one carrier protein over another as substrate by a certain PPTase is probably determined by differences in charge and hydrophobic nature of the helices of PCPs and ACPs. Additionally, the overall charge of the whole carrier protein as well as steric reasons may also play a role in protein-protein recognition [98].

BS168 contains three gene clusters devoted to nonribosomal synthesis of surfactin, fengycin, the siderophore bacillibactin and an unknown polyketide from the *pksX* gene cluster, but is deficient in their synthesis due to a frame shift mutation on the *sfp* gene [99]. Restoring a functional Sfp in such strains led to the production of surfactin, however production of a polyketide by PksX was not detected [65].

1.3.2.4 4'-phosphopantetheinyl transferase independent nonribosomal synthesized antibiotics

Bacilysin is one of the simplest antibiotics, which contains only two amino acids, namely a L-alanine at the N-terminus and a non-proteinogenic amino acid, L-anticapsin, at the C-terminus[100]. This antibiotic is widely spread in different *Bacilli*, like *B. subtilis*, *B. pumilus and B. amyloliquefaciens* [101]. The antibiotically active part of bacilysin is anticapsin that can be released after importing susceptible cells by peptidases and afterwards inhibits the activity of glucosamine synthetase [102]. The gene cluster *bacABCDE* (formerly designated *ywfABCDE*) was found in *B. subtilis* 168, which is involved in the biosynthesis of bacilysin [101]. The mutation and over-expression of these genes showed that BacABC are responsible for the biosynthesis of L-anticapsin, BacD ligates L-alanine with L-anticapsin and BacE seems to be involved in the self protection from bacilysin [101].

1.4 Aim of this work

Although it is known that PGPR enhance plant growth by direct and indirect means (detailed in 1.2), specific mechanisms have not been well characterized, especially for Gram-positive

bacteria like *Bacilli*. Sequencing of the whole genome of *B. amyloliquefaciens* FZB42 allows comparing the genome of the plant growth promotion strain FZB42 with the model strain BS168 to find out genes that could be involved in plant growth promotion. Generating mutants altered in expression of the respective genes helps to understand their precise role in the process of the plant growth promotion, especially in the production of secondary metabolites.

Two dozens of antibiotics produced by *Bacilli* have been already identified and characterized. To find out the capacity in producing secondary metabolites in FZB42, genes involved in secondary metabolism, especially in antibiotics production, are analyzed by using the whole genome sequence of FZB42. Because the majority of these already identified antibiotics are lipopeptides and bacteriocins and only little knowledge exists about the polyketide production in *Bacilli*, analysis of polyketide synthesis is the major objective of this work. Gene clusters involved in polyketide synthesis will be identified and their biosynthesis pathways will be illuminated by using the information from the genome sequence of FZB42 and genetic manipulations. The Ppant transferase Sfp is essential for lipopeptide production in *Bacilli*. This work evidenced for the first time the role of Sfp in biosynthesis of polyketides in *Bacilli*.

We already know that *Bacillus amyloliquefaciens* FZB42 is naturally competent, but the timing and duration of competence in FZB42 are still unclear [8, 16]. The FZB42 cells become competent under laboratory conditions using a minimal medium according to Rapoport´s method [103]. The Spizizen´s protocol was used to obtain competent cells of *B. subtilis* 168 [104]. Therefore, it is difficult to compare the competence event of FZB42 to BS168 in two different medium systems. For this reason, duration and timing of the competence was investigated in this work with both FZB42 and BS168 according to Rapoport´s method [103]. The possible reasons for different competence behaviors between FZB42 and BS168 were analyzed by comparing both genomes.

Many of the results obtained in course of my PhD work have been already published. Therefore, these results are only shortly summarized here and the main space in this thesis is given to the results still unpublished. The papers, in which I contributed, are cited with the tag "Lit":

Lit. 1 **Chen XH**, Koumoutsi A, Scholz R, Borriss R. Genome analysis of Bacillus amyloliquefaciens FZB42 reveals its potential for biocontrol of plant pathogens. *accepted* J Biotechnol. 2009.

Lit. 2 **Chen XH**, Scholz R, Borriss M, Junge H, Mögel G, Kunz S, Borriss R. Dif ficidin and bacilysin produced by plant-associated *Bacillus amylolique faciens* are ef ficient in controlling fire blight disease. *accepted* J. Biotechnol. 2009.

Lit. 3 Moldenhauer J, **Chen XH** , Borriss R, Piel J. Biosynt hesis of the ant ibiotic bacillaene, the product of a giant polyketide synt hase complex of the trans-AT family. Angew Chem Int Ed Engl. 2007; 46(43):8195-7.

Lit. 4 Schneider K*, **Chen XH***, Vater J, Franke P, Nicholson G, Borriss R, Süssmuth RD. Macrolactin is the po lyketide biosynthesi s product of the pks2 cluster of Bacillu s amyloliquefaciens FZB42. J Nat Prod. 2007 Sep; 70(9):1417-23. Epub 2007 Sep 1 1. * These authors contributed equally.

Lit. 5 Koumoutsi A, **Chen XH**, Vater J, Borriss R. DegU and YczE positively regulate the synthesis of bacillom ycin D by Bacillus am yloliquefaciens strain F ZB42. Appl Environ Microbiol. 2007 Nov;73(21):6953-64. Epub 2007 Sep 7.

Lit. 6 **Chen XH** , Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heine meyer I, Morgenstern B, Voss B, Hess WR, Reva O, Junge H, Voigt B, Jungblut PR, Vater J, Süssmuth R, Liesegang H, Strittmatter A, Gottschalk G, Borriss R. Comparative analysis of the complete genome sequence of the plant growth-prom oting bacterium Bacillus a myloliquefaciens FZB42. Nat Biotechnol. 2007 Sep; 25(9):1007-14. Epub 2007 Aug 19.)

Lit. 7 **Chen XH**, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Gra mmel N, S trittmatter A W, Gottschalk G , Süss muth RD, Borriss R. S tructural and functional charac terization of thr ee polyke tide synthas e gene c lusters in Bacillu s amyloliquefaciens FZB 42. J Bacteriol. 2006 Jun; 188(11):4024-36.

Lit. 8 Koumoutsi A, **Chen XH** , Henne A, Liesegang H, Hitzer oth G, Franke P , Vater J, Borriss R. S tructural and functional characteriza tion of gene clusters directing nonribosom al synthesis of bioactive cyclic lipop eptides in Bacillu s am yloliquefaciens strain FZB42. J Bacteriol. 2004 Feb;186(4):1084-96.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Manufacturer	Products
Bio-Rad	Low-Melt agarose gel
Fluka	$CaCl_2$, EDTA
Promega	BCIP (50 mg/ml), NBT (50 mg/ml)
Roth	Agarose, acetonitrile, citric acid, $CuSO_4$, DEPC, $FeCl_2$, $FeCl_3$, $Fe_2(SO_4)_3$, formaldehyde, formic acid, L-glutamic acid, glycerol, IPTG, KCl, K_2HPO_4, H_2KPO_4, maleic acid, methanol, $MgSO_4$, $MnCl_2$, $MnSO_4$, Na-acetate, Na-citrate, Na_2CO_3, NaCl, NaOH, $(NH_4)_2SO_4$, peptone, Pipes, SDS, Proteinase K, Tris, Tween 20, XGal, yeast extract, $ZnCl_2$
Roche	Anti-DIG AP, blocking reagent, DIG-dUTP
Serva	Agar, casamino acids, EGTA, glucose, N-lauroylsarcosine-sodium, $MgCl_2$, MOPS, NaN_3, Na_2SO_4, ONPG, L-tryptophan
Sigma-Aldrich	Chrome azurol S (CAS), hexadecyltrimethylammoniumbromid (HDTMA)

2.1.2 Buffers, mediums and other supplements

a. Generally used buffers:

Buffer	Compositions
6x Loading buffer	2 mM EDTA (pH8.0), 15% Glycerol, 0.1% Bromophenyl blue
10x TAE buffer	400 mM Tris, 200 acetic acid, 1 mM EDTA-Na, pH 8
10x TBE buffer	890 mM TRIS, 890 mM boric acid, 20 mM EDTA-Na, pH 8
10x TE buffer	100mM Tris-Cl (pH = 7.4), 10mM EDTA, pH 7.4 adjusted by acetic acid

b. Generally used mediums:

Medium Compositio ns	
Landy medium	2% glucose, 0.5% Na-glutam ate (pH = 7.4), 0.05% $MgSO_4$,0.05% KCl, 0.1% KH_2PO_4, 15µg% $Fe_2(SO_4)_3.6H_2O$, 500µg% $MnSO_4.H_2O$, 16µg% $CuSO_4.5H_2O$
LB medium	1% pepton, 0.5% yeast extract, 0.5% NaCl, adjust to pH 7
Malt medium	3% malt
Potato medium	4% potato puree (without milk), 2% glucose
SOB medium	0.5% yeast extract, 2% tr ypton, 10 m M NaCl, 2.5 m M KCl, 10 mM $MgCl_2$ 0.5%, 10 mM $MgSO_4$
Wachsman medium	0.5% peptone, 1% glucose, 0.3% yeast extract, 85 mM NaCl

1.5% Agar was added to obtain solid mediums.

c. Supplements used in mediums

Supplement (manufacturer)	**Final concentrations**
Amplicillin, Roth	100 µg/ml
Chloramphenicol, Roth	20 µg/ml (for *E. coli*), 5 µg/ml (for *Bacilli*)
Erythromycin, Serva	1 µg/ml (for Bacilli)
IPTG, Roche	1 mM
Kanamycin, Roche	20 µg/ml (for *E. coli*), 5 µg/ml (for *Bacilli*)
Lincomycin, Serva	25 µg/ml (for Bacilli)
XGal, Roche	40 µg/ml

2.1.3 Enzymes

Vent Polymerase and β-Agarase I was supplied by Ne w England Biolabs. T aKaRa and Eppendorf, repectively, provided TaKaRa LA Taq and Triple Master Enzyme. Taq Polymerase was perchased from Bioron, and dNTPs were from Bioron or Fermentas.

Fermentas or New England Biolabs supplied rest riction endonucleases. CI AP (Calf intestinal alkaline phosphatases), DNA markers, dNTPs, Klenow Fragm ent, RevertAid M-MuL V reverse tran scriptase (2 00U/µl), restriction endonucleases, RiboLock ribonuclease inhibitor (40U/ µl), T4 DNA ligase, T4 Polynucleotide kinase were provided by Fermentas.

2.1.4 Equipments

ABI PRISM model 3730XL	Applied Biosystems, Darmstadt, Germany
ABI PRISM model 377-96	Applied Biosystems, Darmstadt, Germany
Agagel Mini agaros e gel electrophorese chamber	Biometra, Göttingen, Germany
agarose gel electrophorese chamber	Biorad, München, germany
Agilent 1100 HPLC system	Agilent, Waldbronn, Germany
Agilent 1200 HPLC system	Agilent, Waldbronn, Germany
APEX II FTICR mass spectrometer 4.7 T	Bruker-Daltonics, Bremen, Germany
Applied Biosystems QTrap 2000	Applied Biosystems, Darmstadt, Germany
Centrifuge MIKRO 200 Classic Sprinter	Buckinghamshire, GB
Centrifuge MIKRO 200R Cool Sprinter	Buckinghamshire, GB
Centrifuge Sorvall RC 5B; Rotor GS3 & SS34	Du Pont/ Sorvall Instrum ents, Wilmington, US
CHEF Mapper XA System	Bio-Rad, Munich
DRX 500 NMR spectrometer	Bruker, Karlsruhe, Germany
Gene Genius UV imager	Syngene, Cambridge, UK
Gene Quant II photometer	Pharmacia, Jena, Germany
GeneAmp PCR system 2700	Applied Biosystems, Darmstadt, Germany
Luna 3 µ C18 colum n (100 A, 50 × 1.0 mm) for LC-ESI MS	Phenomenex, Aschaffenburg, Germany
Luna 5 µ C18 column (100 A, 100 × 4.6 mm) for preparative HPLC	Phenomenex, Aschaffenburg, Germany
MegaBace1000 sequencers	GE Healthcare
MegaBace4000 sequencers	GE Healthcare
Peltier Thermal Cycler PTC-200	Biozym, Hamburg, Germany

pH meter 761 Calimatic	Knick, Berlin, Germany
rotary evaporator Heidolph VV2000 Heidolph,	Schwabach, Germany
Ultraspec 2000 UV/visible photometer	Pharmacia, Jena, Germany
Vacuum blotter model 785	Bio-Rad, Munich
ZORBAX Eclipse XDB-C18 (5 µ m, 4.6 x 150 mm)	Agilent, Waldbronn, Germany

2.1.5 Kits and others

Manufacturer	**Products**
Amersham	Ready to Go DNA labelled Beads
Epicentre	MasterPure™ DNA & RNA Purification Kit, CopyControl™ fosmid library production kit and Lambda Packaging Extract
Qiagen	QIAEX II gel extraction kit, QIAprep Spin m ini prep kit, Qiaquick PCR purification k it, R.E.A.L. Prep Kit and Polypropylene Column (5 ml)
Macherey-Nagel	Nucleo Spin ® Extract II, Nucleo Spin RNA L, Nucleospin plasmid kit, and Porablot NY plus
Merck	HPTLC aluminium sheet with Sillica gel 60 F_{254}
Millipore	Micron YM-100 Column
Sigma Am	berlite XAD-16
Invitrogen Nebulizer	

2.2 Strains, plasmids and oligonucleotides

2.2.1 Strains

a. Strains used generally in this work:

Strain Genotype		**Reference**
Bacillus amyloliquefaciens FZB 42	Wild type	ABiTEP GmBH, Berlin, Germany
AK1 FZB42	*bmyA*::em^R	[8]
AK2 FZB42	*fenA*::cm^R [8]	

Strain Genotype		**Reference**
AK3 FZB42	*bmyA*::em^R *fenA*::cm^R	[8]
AK15	FZB42 *bmyA-His* $_6$ *-Taq*::*cm*^{R}	[105]
AK36 FZB42	Δ*sigW*::*cm*^{R}	[106]
TF1 FZB42	Δ*degU*::*em*^{R}	Tzu-Fang Huang, unpublished
Bacillus subtilis 168	*trpC2*	Laboratory stock
Bacillus subtilis OKB105	JH642; *pheA1, sfp* [65]	
E. coli DH5α	Δ*lacU169 hsdR17 relA1*	Laboratory stock
E. coli Stbl2	JM109; *ΔrecA1, ΔmcrA*	Invitrogen, Karlsruhe, Germany
E. coli GM2163	*K-12; Δdam, Δdcm*	Fermentas, St.Leon-Rot, Germany

b. Strains as indicators for antibiotics screening:

Strain Genotype		**Reference**
Fusarium oxysporum	Wild type	Laboratory stock
Saccharomyces cerevisiae	Wild type	Laboratory stock
Bacillus megaterium	Wild type	Laboratory stock

c. Strains constructed in this work:

Strain Genotype	
CH1 FZB42	Δ*srfAA*::*em*^{R}
CH2 FZB42	*fenA*::cm^R Δ*srfA*::*em*^{R}
CH3 FZB42	Δ*sfp*::*em*^{R}
CH4 FZB42	Δ*yczE*::*em*^{R}
CH5 FZB42	Δ*sfpyczE*::*cm*^{R}
CH6 FZB42	Δ*bae*::*cm*^{R}

Strain Genotype	
CH7 FZB42	Δ*mln::cm*^{R}
CH8 FZB42	Δ*dfn::em*^{R}
CH9 FZB42	Δ*dfnAT::em*^{R}
CH10 FZB42	Δ*dfnTE::cm*^{R}
CH11 FZB42	Δ*bae::cm*^{R} Δ*dfn::em*^{R}
CH12 FZB42	Δ*mln::cm*^{R} Δ*dfn::em*^{R}
CH13 FZB42	Δ*mln::neo*^{R}
CH14 FZB42	Δ*bae::cm*^{R} Δ*mln::neo*^{R}
CH15 FZB42	Δ *dfnK::cm*^{R}
CH16 FZB42	Δ *dfnL::em*^{R}
CH17 FZB42	Δ *dfnM::em*^{R}
CH18 FZB42	Δ*bmyA*::em^R Δ*mln::neo*^{R}
CH19 FZB42	Δ*bmyA*::em^R Δ*bae::cm*^{R} Δ*mln::neo*^{R}
CH20 FZB42	Δ*bmyA-His* $_6$ *-Taq*::*neom*^{R}
CH21 FZB42	Δ*bae::cm*^{R} Δ*dfn::em*^{R} Δ*bmyA-His* $_6$ *-Taq*::*neom*^{R}
CH22 FZB42	Δ*mln::cm*^{R} Δ*dfn::em*^{R} Δ*bmyD::cm*^{R}
CH23 FZB42	Δ*comA*::em^R
CH24 FZB42	Δ*bae1ATcmR::*
CH25 FZB42	Δ*bae2AT*::em^R
CH26 FZB42	Δ*bae3AT*::em^R
CH27 FZB42	Δ*mlnAT*::em^R
CH28 FZB42	Δ*pksA*::em^R
CH29 FZB42	Δ*ganA*::em^R
CH30 FZB42	Δ*sigV*::em^R
CH31 FZB42	Δ*ganA::cm*^{R}
CH32 FZB42	Δ*sigB*::em^R
CH33 FZB42	Δ*bae::neo*^{R} Δ*dfn::em*^{R}
CH34 FZB42	Δ*mln::neom*^{R} Δ*dfn::em*^{R}

Strain Genotype	
CH35 FZB42	Δ*sigV*::em^R Δ*sigW*::cm^R
CH36 FZB42	Δ*nrsC::cm*^{R}
CH37 FZB42	Δ*nrs::cm*^{R}
CH38 FZB42	Δ*dhb*::em^R
CH39 FZB42	Δ*dhb*::em^R
CH40	*lacI cm*^{R}*, pSpac* as promotor for the *bmy* gene cluster
CH41	*lacI cm*^{R}*, pSpac* as promotor for *bmy* gene cluster*,* Δ*yczE::em*^{R}
CH42	*lacI cm*^{R}*, pSpac* as promotor for *bmy* gene cluster, Δ*degU::em*^{R}

d. Strains for antibiotics screening

Strain Refer	**ence**
Bacillus amyloliquefaciens BE 20-2	laboratory stock
Bacillus subtilis subtilis INR 937	Joseph Kloepper, Auburn, AL, US
Bacillus subtilis subtilis FZB 37	ABiTEP GmBH, Berlin, Germany
Bacillus subtilis spizenii UCMB-5014	Oleg Reva, Ukranian Collection of Microorganisms, Kyiv, Ukraine
Bacillus subtilis spizenii At3	Oleg Reva, Ukranian Collection of Microorganisms, Kyiv, Ukraine
Bacillus amyloliquefaciens UCMB-5044	Oleg Reva, Ukranian Collection of Microorganisms, Kyiv, Ukraine
Bacillus amyloliquefaciens QST713 (Serenade)	Serenade, Agraquest Inc., Davis, CA, US
Bacillus amyloliquefaciens FDK21	ABiTEP GmBH, Berlin, Germany
Bacillus amyloliquefaciens B9601-Y2	He, Yueqiu, Yunnan agriculture university, China
Bacillus amyloliquefaciens UCMB-5113	Oleg Reva, Ukranian Collection of Microorganisms, Kyiv, Ukraine
Bacillus amyloliquefaciens INR-7	Joseph Kloepper, Auburn, AL, US
Bacillus amyloliquefaciens FZB 24	ABiTEP GmBH, Berlin, Germany

Strain	**Reference**
Bacillus amyloliquefaciens FZB 113	ABiTEP GmBH, Berlin, Germany
Bacillus amyloliquefaciens FZB 111 (Kodiak)	ABiTEP GmBH, Berlin, Germany
Bacillus amyloliquefaciens FZB 109 (A1/3)	ABiTEP GmBH, Berlin, Germany
Bacillus amyloliquefaciens GB 03 (Kodiak)	Kodiak, Gustafson Biologicals, Plano, TX, US
Bacillus amyloliquefaciens DSM7	laboratory stock
Bacilus licheniformis S175 (NCIMB 7724)	Fergus Priest, Edinburgh, Great Britain
Bacilus licheniformis ATCC 9789 (B27)	Fergus Priest, Edinburgh, Great Britain
Bacilus licheniformis S167 (9945A)	Fergus Priest, Edinburgh, Great Britain
Bacilus licheniformis S173 (Gibson 1158)	Fergus Priest, Edinburgh, Great Britain
Bacilus licheniformis (B37)	laboratory stock
Bacillus amyloliquefaciens1 FZB 45	ABiTEP GmBH, Berlin, Germany
Bacillus mojavensis UCMB-5075	Oleg Reva, Ukranian Collection of Microorganisms, Kyiv, Ukraine
Bacillus mojavensis MB2	ABiTEP GmBH, Berlin, Germany
Bacillus amyloliquefaciens GaoB2	Gao, Xuewen, Nanjing agriculture university, China
Bacillus amyloliquefaciens GaoB3	Gao, Xuewen, Nanjing agriculture university, China
Bacillus amyloliquefaciens Gao55	Gao, Xuewen, Nanjing agriculture university, China
Bacillus amyloliquefaciens Gao85	Gao, Xuewen, Nanjing agriculture university, China

2.2.2 Plasmids

a. Genernally used plasmids:

plasmid description	
pDG148 [107]	*E. coli* and *B. subtilis* shuttle vector, IPTG-inducible P_{spac} promoter, Ap^R Km^R
pDG268 [108]	pBR322 de rivarive with promoterless *lacZ*, integrative vector for recombination into *B. subtilis amyE*, Ap^R Cm^R
pMX39 [109]	*E. coli* and *B. subtilis* shuttle vector based on pBR322 and PDB101, Ap^R Em^R
ECE73 [110]	Cm^R to $Neom^R$ exchange vector, Ap^R
ECE149 [111]	Protein fusion vector for *Bacillus*, Gfp with IPTG-inducible P_{spac} promoter, Ap^R Em^R
pGEM-T (Promega, Mannheim)	Cloning vector, Ap^R
pCR2.2 TOPO (Invitrogen, Karlsruhe)	Cloning vector, Ap^R
pTZ19R (GE Healthcare, Munich)	Cloning vector, Ap^R

b. Plasmids constructed in this work:

plasmid	**Description**
pGEM-TcmR	992 bp PCR fragment containing *cm*^{R} was ligated in pGEM-T
pUCemR	1791 bp fragm ent containing *em*^{R} from pMX39 cut by HindIII, inse rting in pUC18 at HindIII
pUCcmR	1861 bp fragm ent containing *cm*^{R} from pDG268 cut by SphI and HindIII, inserting in pUC18 at SalI and HindIII
pCH01a	2285 bp PCR fragment containing *srfAA* was ligated in pGEM-T
PCH01	1798 bp containing *em*^{R} from pUCemR cut by HindIII was inserted in pCH01a cut by HindIII
pCH03a	2281 bp fragment containing *sfp* was ligated in pGEM-T
pCH03	1152 bp containing *em*^{R} from pU CemR w as inserted in pCH03a cut by MunI and Kpn2I (101 bp of *sfp* was deleted)
pCH04a	2517 bp fragment containing *yczE* was ligated in pGEM-T
pCH04	1152 bp containing *em*^{R} from pUCemR was inserted in pCH04a cut by Sm aI and EcoRI (386 bp of *yczE* was deleted)
pCH05a	2090 bp fragment containing *sfp* and *yczE* was ligated in pGEM-T
pCH05	1333 bp containing *cm*^{R} from pUCcmR was inserted in pCH05a cut by Hindlll and Aval (783 bp of *sfp* and *yczE* was deleted)
pCH06a	3318 bp fragment containing KS2 of *baeJ* was ligated in pGEM-T
pCH06	1333 bp containing *cm*^{R} from pUCcmR was inserted in pC H06a cut by EcoRI and Hpal (1360 bp of *baeJ* was deleted)
pCH07a	2984 bp fragment containing mlnBKS0 was ligated in pGEM-T
pCH07	1333 bp containing *cm*^{R} from pUCcmR was inserted in pCH07a cut by EcoRl and Kpnl (1257 bp was deleted)
pCH08a	2128 bp fragment containing KS1of *dfnD* was ligated in pGEM-T
pCH08	1791 bp containing *em*^{R} from pUCemR was inserted in pCH08a cut by EcoRl and EcoR72l (376 bp of *dfnD* was deleted)
pCH09a	2973 bp fragment containing *dfnAT* was ligated in pGEM-T

plasmid	Description
pCH09	1524 bp containing *em*^{R} from pUCemR was inserted in pC H09a cut tw ice by Aval (781 bp of *dfnAT* was deleted)
pCH10a	2565 bp fra gment containing TE dom ain of the *dfn* gene clu ster was lig ated in pGEM-T
pCH10	1345 bp containing *cm*^{R} from pUCc mR was inserted in pCH10a cut by EcoRl and BamHl (523 bp of TE domain was deleted)
PCH15a	1498 bp fragment containing *dfnK* was ligated in pGEM-T
pCH15	992 bp containing *cm*^{R} from pUCcmR was inserted in pCH15a cut by Kpnl and Eco91l (771 bp of *dfnK* was deleted)
pCH16a	2076 bp fragment containing *dfnLM* was ligated in pGEM-T
pCH16	1126 bp PCR fragm ent containing emR was inserted in pC H16a cut by Hindlll and BamHl (28 bp of *dfnL* was deleted)
pCH17	1220 bp containing *em*^{R} from pUCemR was inserted in pC H16a cut by Kpn2 and Bpu1102l (197 bp of *dfnM* was deleted)
pCH23a	2092 bp fragment containing *comA* was ligated in pGEM-T
pCH23	1152 bp containing *em*^{R} from pUCemR was ins erted in pCH23a cut by KspAI and Eco72I (104 bp was deleted)
pCH24a	2741 bp fragment containing *baeAT1* was ligated in pGEM-T
pCH24	1333 bp containing *cm*^{R} from pUCcmR was inserted in pC H24a cut by HindIII and HpaI (462 bp of *baeAT1* was deleted)
pCH25a	2741 bp fragment containing *baeAT2* was ligated in pGEM-T
pCH25 1	152 bp containing *em*^{R} from pUCemR was inserted in pCH25a cut by HindIII and Eco47III (928 bp of *baeAT2* was deleted)
pCH26a	1401 bp PCR fragm ent cut by NdeI a nd NaeI w as inserted in pUCe mR cut by NdeI and EheI
pCH26	1842 bp PCR fragm ent cut by SphI was inserted in pCH26a cut by HhpaI and SphI (193 bp of the *baeAT3* was deleted)
pCH27a	2466 bp fragment containing *mlnAT* was ligated in pGEM-T

plasmid	Description
pCH27 1	152 bp containing *em*^{R} from pUCemR was inserted in pCH27a cut by HindIII and EcoRI (472 bp of *mlnAT* was deleted)
pCH28a	2532 bp fragment containing *baeA* was ligated in pGEM-T
pCH28	1152 bp PCR fragm ent containing *em*^{R} was inserted in pCH28a cut twice by BgllI (306 bp of *baeA* was deleted)
pCH29a	1971 bp fragment containing *ganA* was ligated in pGEM-T
pCH29	1152 bp containing *em*^{R} from pUCemR was inserted in pCH29a cut by SnaBI
pCH30a	1286 bp PCR fragment cut by EcoRI was inserted in pUCemR cut by HpaI und EcoRI
pCH30	1892 bp PCR fragm ent of cut by PvuII und NdeII was inserted in pCH30a cut by NdeI and EheI (*sigV* was deleted)
pCH31	992 bp containing *cm*^{R} from pUCcmR was inserted in pCH29a cut by SnaBI
pCH32a	3064 bp PCR fragm ent cut by NdeI a nd SmiI was inserted in pUCe mR cut by NdeI and EheI
pCH32	3064 bp PCR fragm ent of cut by Ps tI was inserted in pUCemR cut by PstI and SmaI (*sigB* was deleted)
pCH37a	2742 bp fragment containing *nrsC* was ligated in pGEM-T
pCH37	1232 bp containing *cm*^{R} from pUCcmR was inserted in pC H37a cut tw ice by EcoR and HindIII (706 bp of *nrsC* was deleted)
pCH38a	1360 bp PCR fragm ent cut by Bam HI was inserted in pUCcmR cut by Ba mHI and NaeI
pCH38	2167 bp PCR fragm ent cut by Hind III was inserted in pCH38a cut by HindIII and EheI (deletion of whole *nrs* gene cluster)
pCH39a	1993 bp PCR fragm ent cut by EheI a nd NdeI was inserted in pUCe mR cut by EheI and NdeI
pCH39	13045 bp PCR fragment of cut by HpaI and PstI was inserted in pCH39a cut by PstI and PvuII (deletion of whole *dhb* gene cluster)
pCH41a	1296 bp containing LacI of ECE149/BamHI and SmiI was inserted in pUCcmR cut by Eco47III and BamHI

plasmid	Description
pCH41b	1322 bp PCR fragment was inserted in pCH41 between EcoRI and HindIII cutting site
pCH41c	1536 bp PCR fragment cut by SacI and BamhI was inserted in pCH41b cut by SacI and BamHI, exchange P_{bmy} with P_{spac}

2.2.3 Oligonucleotides

Over 900 primers were designed during *Bacillus amyloliquefaciens* FZB42 sequencing project. 16 primers were designed for long range PCR, and 38 primers were designed for multiplex PCR (see lists below). Oligonucleotides were provided from Sigma Aldrich, Operon or Eurogentec.

a. Oligonucleotides used in genomic project for long range PCR

Primer	Sequence 5´-> 3´	Contig name
babg57x1993	AATGTTAATGCGACTTCGCCCGCTTTCGTCTCC	gbacg30f
back95x1995	TGCCTCTGCTCTCAGGCTTCCTTATCAACAACCAC	abald13r
babg57x1993	AATGTTAATGCGACTTCGCCCGCTTTCGTCTCC	gbacg30f
baey96x1996	AGCGTCAATCCGACATTCCCCTTATACATCGCAGC	abald13r
bajh39x1994	GATGTCCGTCTTTTACCATACGTTCAACCGTGTCCG	gbacg30f
back95x1995	TGCCTCTGCTCTCAGGCTTCCTTATCAACAACCAC	abald13r
bajh39x1994	GATGTCCGTCTTTTACCATACGTTCAACCGTGTCCG	gbacg30f
baey96x1996	AGCGTCAATCCGACATTCCCCTTATACATCGCAGC	abald13r
badu60x1840	TGCTCCGTCAGACTTTCGTCCATTG	on 16S rRNA
baqi14x1876	AACAACAGAGCTTTACGATCCG	on 16S rRNA
baqi14x1873	TATCTAATCCTGTTCGCTCCCC	on 16S rRNA
baqi14x1855	AGAGTGCCCAACTGAATGCTGGCAAC	on 16S rRNA
badu60x1840	TGCTCCGTCAGACTTTCGTCCATTG	on 16S rRNA
baqi14x1876	AACAACAGAGCTTTACGATCCG	on 16S rRNA
baqi14x1873	TATCTAATCCTGTTCGCTCCCC	on 16S rRNA
baqi14x1855	AGAGTGCCCAACTGAATGCTGGCAAC	on 16S rRNA

b. Oligonucleotides used for multiplex PCR

Primer	Sequence 5´-> 3´	Contig name
badh62x1859	GTACTGGGTTTGAAGGCACTCG	abald13r
bafk72x1953	TCAGCGGCAAGGACATTCCTTATGACAGAC	aba1876pcr0960
bajo94x2059	TTTGGCTTTGCCACATCCCTAC	aba1954pcr0872
bagf09x1810	ACACTGGATGAGCTGACATACG	ababy33x1809w2
baby33x1966	ATCATCGGCAATGTGCTTGGTG	ababy33x1809w2
baeb27x2085	AACGAGATCGGCGAAAACAGC	abaeb27f
baeb27x2084	TGGACGAACGCCAGAAGATCAC	abaeb27f
baaw06x2082	AACCATTAACCTGTGCCAGCCC	gbaaw06f
baaw06x2083	ATCGCCATTACACCGCAGATTC	gbaaw06f
baaw25x2074	TAATAACGCTGCACCCACTGCC	gbaaw25f
baaw25x2075	TGCTGGCGACCAAAATCACCAC	gbaaw25f
baji93x1860	ACATCGCACGCTTCAGCTTCAC	gbafh19f
badv31x2036	TTCGTTCAGCCTTTGCCGTCAC	gbafh19f
bafy85x2073	TCCCTCAGCGGTTAGCAAGAAG	gbafy85f
bafy85x2072	TTTACCCGTAACGACTGACCCC	gbafy85f
banc82x2062	ACTCATGGGGTGTTCGTACATC	gbagm27r
bajg64x2061	TGAATGAAGAGAAGGTGTCCCG	gbagm27r
baid72x2066	AACTTCGACCTCCATACAAGCC	gbaid72f
baid72x2067	ACGTGGTCACCCTTCTTTACGC	gbaid72f
babt35x0767	ACCTATTGGACGTTAAAGGAC	gbakj78r
babp88x2063	TCGGTTATGACGGCACTGAAAC	gbakj78r
bakk20x2078	TCCACTTTTGTGCTTCGTGCCC	gbakk20f
bakk20x2079	TGTTCTGAGGAGTCCCTGATGG	gbakk20f
baks42x2068	AGAGTAGGGCTTCGGGTGTTTG	gbaks42f
baks42x2069	TGTTGAACGTGTCGTTACCGTG	gbaks42f
bakt57x2064	TTCAACTGCTTGCCCGTAATGG	gbakt57f
bakt80x2065	TTTCGGAACTTGAGGATGTCGC	gbakt57f

Primer	Sequence 5´-> 3´	Contig name
baah04x1961	AGAGGTGGATACATAATGCCGAC	gbakv75r
balm29x2071	TTGTGCCGACCTGGTACGAATC	gbalm29f
balm29x2070	TATACCGACGAAGCCGCATTGG	gbalm29f
bacj30x2077	AAACTGCTGCGGAATAACGCTC	gbanf96f
banf23x2076	TTTCATCGGAATCTGCGGCATC	gbanf96f
bagh57x2060	AGGACGATGTTGGAAGTAGAG	gbanz86f
banz86x1967	TCAATGACTTCAAGCGGCTGGC	gbanz86f
bapr63x2080	ATTCACACTGGCACGGACAAGG	gbapr63f
bapr63x2081	TATACCTTTGCTTCCACGGCAG	gbapr63f
baqs37x1988	TGCAAACAGCCTAGATAAAGATATATGGTG	gbaqs37f

c. Oligonucleotides designed for secondary metabolites

Primer Sequence	5´→ 3´	Applications
srfAAkn3-1	AGCCGTCCTGTCTGACGACG	pCH01a
srfAAkn3-2	GACTATGCGGTATGGCAGCAGA	pCH01a
sfp-n1	ATCTGATGGTGGAGCAGCTTG	pCH03a
sfp-n2	TTGTAGGAGCGGGAGAAGAAGC	pCH03a
yczE-n1	ACTCAACGTGTCCAACGTCAAG	pCH04a
yczE-n2	AAGGATACATCTCTCGCGTCAC	pCH04a
sfp1	TCAACGTGTCCAACGTCAAG	pCH05a
ycz2	TTGGCGGTTATGCTACAATG	pCH05a
pks1Up	AAAGGAGCGGAGTGCAACATC	pCH06a
pks1Dw	TGAGATGATGCCGTCCTCTTC	pCH06a
pks2Up	AGCTCATTGACAGCGATGCTGC	pCH07a
pks2Dw	ATGATCGCCGCTTCCTCATCAG	pCH07a
pks3Up	ACATTGACCGCATCCTCAATCTG	pCH08a
pks3Dw	TCAGCTGCTGTTCGGAATGTG	pCH08a

Primer Sequence	5′→ 3′	Applications
pks3AT-1	TGTGATTTTGCTTCATGGACGC	pCH09a
pks3AT-2	TTTCGCATAACCGTGCTTCTCC	pCH09a
pks3TE-1	TTGCAGAGGGTATGATCTTCC	pCH10a
pks3TE-2	TGCATGAAGTATCCGATCATC	pCH10a
pks3Enoyl-1	TCTCAGTCTCAAGGAGTTCCC	PCH15a
pks3Enoyl-2	TTCTGGAAACAATCGGAGAAG	PCH15a
pks3cyp-1	ACCGTTGAGTCAATCGTCATGC	pCH16a
pks3cyp-2	TATCCGTTCCGATGATGAGCG	pCH16a
comA-1	TCTGATGGGATCGCTGATCTG	pCH23a
comA-2	TCTGCATGATTCAGTGCTGCAG	pCH23a
pks1AT1-1	ACAGAAGGGTGCGGTATATGTGGG	pCH24a
pks1AT1-2	TGAGGGCATTGACGACATAGAGC	pCH24a
pks1AT2-1	AGCAGGCTGATGAAACACTCGG	pCH25a
pks1AT2-2	TCTGTTATCGGCTCCGCTTCAG	pCH25a
pks1AT3-1	TCGTCTGTCTGATCCATTTGACCG	pCH26a
pks1AT3-2	TCCTTCGCTTCCAGCATGTATCTG	pCH26a
pks1AT3-3	ATGGCAGCGGTGATCGGCCTG	pCH26b
pks1AT3-4	TCAGAAGCGAGGCCGTCTCTTCC	pCH26b
pks2AT-1	TCTGAGCAATGCAAATCAGCG	pCH27a
pks2AT-2	TGTGTGTTTCCATCATTTCATCCC	pCH27a
pksR-1	AGCTGAGAATGCACATCACGGC	pCH28a
pksR-2	AAGCGTTCCTGACCCGCTCC	pCH28a
araB-1	AGACGAAACCGTCAGCATCATC	pCH29a
araB-2	TCCCTGAAATCCGTGAGGAGC	pCH29a
sigVL-1	AGCTAAAGCGCAGCCGAGAAG	pCH30
sigVL-2	AAGAAATCGCAGACATCACGGG	pCH30
sigVR-1	TCCTTGCGCTCAGTTATGCAC	pCH30a
sigVR-2	AATGAAAGCGGCAAGATGGCAC	pCH30a

Primer Sequence	**5′→ 3′**	**Applications**
sigB-1	AGAACGCAGCGACATTCAGCG	pCH33a, pCH33
sigB-2	TCTCGTCATCTGCGGCTGACTTG	pCH33a, pCH33
nrs-1	TTTGCTAGTTTCTCAATGGTAC	CH37a
nrs-2	AGCAGATGGAGCTAGTATCCAG	CH37a
nrsWR-1	TCTGTCGGTGTGCTTACATTGC	CH38a
nrsWRbamHI-2	ACTCGACCTAGGATCCAGCATCTTAGGTAAGGCG	CH38a
nrswL-1	TTCATCAGCTCATTGGCGTAAG	pCH38
nrswL-2	TGAGGCTACAGTTGCTGTTAGTG	pCH38
dhbR-1	AATCATGTACCGGCTTACAGC	CH39a
dhbR-2	TATCAAAGCCTTTCCTGAAGCC	CH39a
dhbL-1	AGGACAGAAAGCAAGTATGGG	CH39
dhbL-2	AAATGCTGTGTGAGCTGTTCG	CH39
bmyDR-1-HindIII	AAAGTAAAGCTTCTATTGAGCGGTATACCCG	pCH41a
bmyDR-2	AGGCAAAGTGACGATTGCAGAGG	pCH41a
pspac-1-BamHI	TTTGAGGATCCCCTTGCCTACCTAGC	pCH41b/SOE PCR
pspac-2	TCCTCCGATCAATAGCGCTGGTACCCTC	pCH41b/SOE PCR
bmyDL-1	ACCAGCGCTATTGATCGGAGGAATCTCATG	pCH41b/SOE PCR
bmyDL-2-sacI	TTATCCCGAGCTCAGAGATATTGCTTC	pCH41b/SOE PCR
pks2AT-3	TGTGGTGCTCATCCAGTCCGATC	check pCH25 and CH25
pks3-1up	ACTCTTGGCGTGACGTCTCTC	check CH08 , CH11, CH12 and pCH08
pks3-1dw	TGAGGAAACCTTCTTCGTGTG	check CH08, CH11, CH12 and pCH08
pks2-1dw	AATCATGGAGAAACAGACATG	check pCH07 and CH07, CH12,
dhbL-M	TGGCCTTCAGCATTGACGAG	check pCH39, CH39 and CH40

Primer Sequence	**5′→ 3′**	**Applications**
dhbR-M	AGGAGGCCGTACGTATCGTATC	check pCH39a, pCH39, CH39 and CH40
bmyDR-M	ACACTCAGCTTGTCCGCAATC	check pCH41b, pCH41c, pCH41, CH41, CH42, CH42
baeC-ATc-1	TCCGTATTCCAACCTTCACAAG	Examining the conserved AT domain in baeC
baeC-ATc-2	AGATAGCGGATGCTGTCGTTCC	Examining the conserved AT domain in baeC
baeD-ATc-1	TCGTCTGTCTGATCCATTTGAC	Examining the conserved AT domain in baeD
baeD-ATc-2	TATATCGGTCTCAGCCTCTCC	Examining the conserved AT domain in baeD
baeE-ATc-1	AGGCTCGCAAAGAATAGGGATG	Examining the conserved AT domain in baeE
baeE-ATc-2	TGCCGTCACATTTGAAATGACC	Examining the conserved AT domain in baeE
pks2AT-c1/mln-c1	ATGCTGTTGCAGGACATAGTC	Examining the conserved AT domain in mlnA
pks2AT-c2/mln-c2	TAGTCAGAATGTTTCCAGGACC	Examining the conserved AT domain in mlnA

Primer Sequence	**5′→ 3′**	**Applications**
difA-ATc-1	AGATCGACATGCCTGTTTCTCC	Examining the conserved AT domain in dfnA
difA-ATc-2	AATGCCTGCCATGATCAGACTC	Examining the conserved AT domain in dfnA
nrs-1f	TGCTCTTCCTTGACACTCAATC	examining the repeats in nrs gene cluster
nrs-1r	AGAAAGAGACAGAACATCACATC	examining the repeats in nrs gene cluster
nrs-2f	TGGTATGCTAACATATAATCTC	examining the repeats in nrs gene cluster
nrs-2r	TCAAGATGTGTTATATGCGTAC	examining the repeats in nrs gene cluster
nrs-PCP2	TCTCTCTCTAAACCCAGGTTGC	examining the repeats in nrs gene cluster
nrs-CD2	AGTGTTGTCGAAGTGGAGTAACC	examining the repeats in nrs gene cluster
nrs-PCP1	TTCCATTCGGAAGAATGACCG	examining the repeats in nrs gene cluster
nrs-CD1	TTCCTTCAACACCAGCTCTGC	examining the repeats in nrs gene cluster
nrs-PCP2	CTCTCTAAACCCAGGTTGC	examining the repeats in nrs gene cluster
nrs-CD2	AGTGTTGTCGAAGTGGAGTAACC	examining the repeats in nrs gene cluster
nrs-PCP1	TTCCATTCGGAAGAATGACCG	examining the repeats in nrs

Primer	Sequence 5′→ 3′	Applications
nrs-CD1	TTCCTTCAACACCAGCTCTGC	examining the repeats in nrs gene cluster
nrsCwhole-1	TTCTGGGTATCCTGCATCCAATAATGCATG	examining the repeats in nrs gene cluster by long range PCR
nrsCwhole-2	AGCTGTGATAGAGCGTCAACACATGCTACG	examining the repeats in nrs gene cluster by long range PCR
nsrSN-1	AATGCTGGTCCATTCAGCCTTG	Southern blot
nsrSN-2	TCGAGACAATCGCCAGTATCC	Southern blot
cmRup	AATAACCTAACTCTCCGTCG	check plasmids and mutant containing cmR
emRdw	TGGGTCAATCGAGAATATCGTC	check plasmids and mutant containing emR
kmRup	AGAGAGCCATAAACACCAATAGCC	check plasmids and mutant containing NeoR
cmBstEII-3	ATGTTTGACAGGTAACCATCGGCAATAGTTACCC	pGEMTcmR
cmKpnl-3	ACCGACGGTACCAAGTACAGTCGGCATTATCTC	pGEMTcmR
emBamHI	ATCATGAGTAGGATCCGAGAGTGATTGGTC	pCH16
emHindIII	TTAATTGGTAAAGCTTAGAGGTGCCCTATACC	pCH16
emRbglII-1	TTAGAAGAGATCTTAAGAGTGTGTTGATAGTGC	pCH28
emRbglII-2	TAATTGAGATCTATTAGAGGTGCCCTATACC	pCH28

The restriction enzyme recognition site within each primer is underlined.

2.3 Methods

2.3.1 Cell biological methods

2.3.1.1 Preparation of calcium chloride competent E. coli *and transformation*

The calcium chloride method was used in this work for preparation of competent *E. coli* according to Sambrook [112].

100 ml LB containing fresh *E. coli* cells was shaken at 37°C and 210 rpm, until O.D. $_{600}$ reached 0.3 to 0.5. The culture was chilled on ice for 10 min, and was washed with 25 ml cold $MgCl_2$ (0.1 M). Cell pellets were resuspended in 5 ml cold $CaCl_2$ (0.1 M), and kept on ice for 20 min. After centrifugation under 2500 xg at 4°C for 10 min, the pellet was resuspended in a mixture with 4.3 ml $CaCl_2$ (0.1M) and 0.7 ml glycerol (87 %). The aliquots (200µl) were stored at - 80°C.

For transformation, an aliquot was thawed on ice for 10 min. 10- to 30-ng plasmid was added in the competent cells. The cells were kept on ice for 45 min, and then a heat shock was performed at 42°C for 30 sec. After adding 800µl LB or SOB (for *E. coli* Stbl2), the culture was shaken under 210 rpm at 37°C or 30°C (for *E. coli* Stbl2) for 1 hour and spread on plates with appropriate antibiotics and other selective compounds.

2.3.1.2 Preparation of competent Bacillus subtilis *and transformation*

According to the method developed by Spizizen [104], fresh cells (not older than 3 days) were grown overnight in 10 ml KM1 buffer at 32°C and 150 rpm, and then 1:10 diluted in 50 ml KM1 buffer. Cells were grown at 37°C and 175 rpm. The cell density was observed every 30 min at wavelength 600 nm. The culture was diluted 1:10 in 100 ml KM2 buffer during the transient phase. Incubation at 37°C and 75 rpm for further 75 min was necessary. Subsequently, the cells were harvested by 10-minute centrifugation at 5000 rpm and room temperature, and the pellet was resuspended in 2-ml supernatant containing 10% glycerol. 0.5 ml aliquots of the competent cell were stored at - 80°C.

For transformation, an aliquot was thawed by short incubation at 37°C. 0.5 ng to 1 µg of the desired DNA (chromosomal or linearized/circular plasmid DNA) was added, and cells were incubated at 37°C for 30 minutes at 50 rpm. Subsequently, 0.5 ml of LB medium, containing sub-lethal concentration (1/10 of the normally used concentration) of the appropriate antibiotic was added to the cells and they were further grown at 37°C and 200 rpm for 75 minutes. The culture was spread on agar plates with the appropriate antibiotics.

Buffer Compositions	
KM1	1X SMM, 1 mM $MgSO_4$, 0.5% glucose, 0.04% casamino acid, 0.005% tryptophan
KM2	1X SMM, 1 mM $MgSO_4$, 0.3% glucose, 0.005% casamino acid, 0.005% tryptophan, 0.001 $CaCl_2$
SMM (10x)	150 mM $(NH_4)_2SO_4$, 800 mM K_2HPO_4, 450 mM KH_2PO_4, 34 mM Na_3citrate

2.3.1.3 Preparation of competent Bacillus amyloliquefaciens *FZB42 and transformation*

Competent cell preparation for *B. amyloliquefaciens* FZB42 was modified base on previously described [103]. Fresh cells were grown overnight in LB medium at 28°C and 180 rpm. The cells were 1:10 diluted in 10 ml GCHE medium and grown at 200 rpm and 37°C. When the $O.D._{600}$ value was near to 1.2, an equal volume of GC medium was added to the culture and grown for another 1 hour at 37°C under 210 rpm. After that, the culture was divided into 2 ml tubes, centrifuged at 6000 rpm and room temperature for 5 min. 100 µl of the supernatant were added into the competent cell pellet and resuspended gently. 1 to 2 µg linearized plasmid DNA were added into the tube and mixed gently for 1 to 2 min. After adding 1 ml at 37°C pre-warmed transformation buffer, the culture was shaken from time to time under 75 rpm at 37°C for 20 min. Before shaking at 37°C and 210 rpm for 90 min, 0.5 ml at 37°C pre-warmed LB medium with a sub-lethal concentration of antibiotics (1/10 of the antibiotic concentration used for *Bacillus*) was added. Finally the culture was spread on LB agar plates with appropriate antibiotics.

Buffer Compositio	**ns**
GCHE	1x PC, 2% glucose, 0.05% tryptophan, 40 mM $FeCl_3$/Na_3citrat, 3 mM $MgSO_4$, 0.1% casein hydrolysat, 0.25% sodium glutamate
GC	1x PC, 2% glucose, 0.05% tryptophan, 40 mM $FeCl_3$/Na_3citrat, 3 mM $MgSO_4$
Transformation buffer	1x SMM, 0.5% glucose, 1 mM EGTA, 20 mM $MgCl_2$
PC (10x)	800 mM K_2HPO_4, 450 mM KH_2PO_4, 28 mM Na_3citrate
SMM (10x)	150 mM $(NH_4)_2SO_4$, 800 mM K_2HPO_4, 450 mM KH_2PO_4, 34 mM Na_3citrate

2.3.1.4 Agar diffusion test

For antifungal tests, *B. amyloliquefaciens* FZB42 and its derivatives were grown in Landy medium at 37°C and 210 rpm for 24 hours. The density of the cultures was adjusted to 0.4 at $O.D._{600}$. 2 µl of these cultures were spotted on Waksman agar together with regularly arranged growing fungi. The plates were incubated at 27°C.

For the antibacterial tests, the indicator strains were grown at 37°C under vigorous shaking until their $O.D._{600}$ reached 0.6. 300 µl of the culture was mixed with 3 ml of LB-agar and poured on dishes. *B. amyloliquefaciens* FZB42 and its derivatives were grown in Landy medium for 36 hours at 30°C and 210 rpm. 20 to 30 µl of their supernatants were added in

the holes on the plates with indicator strains. The plates were incubated at the appropriate temperature.

2.3.1.5 Feeding experiment

Mutant strain CH11 (*Δbae, Δdfn*) was used to produce macrolactin in order to eliminate the background of the other polyketides, like bacillaene and difficidin. The pre-culture was grown in LB broth at 28°C and 180 rpm overnight, and was then 1:100 diluted in Landy medium. To obtain a high uptake of 1-^{13}C sodium acetate, the Landy medium was modified by using a low concentration of glucose (0.25%). The main culture was shaken at 180 rpm and 28°C. 10 mg/ml 1- ^{13}C sodium acetate was added every hour from the 7th to the 12th hour . The supernatant was harvested after further 12-hour cultivation. The investigated compound was purified by XAD16 and isolated by preparative HPLC. Its structure was determined by NMR-spectrometry [113].

2.3.2 Molecular biological methods

DNA digestion with restriction endonucleases, 5'-phosphorylation of oligonucleotide or filling recessed 5'-termini of double- stranded DNA with T4 polynucleotide kinase, filling recessed 3'-termini of double-stranded DNA by Klenow Fragment, dephosphorylation of DNA with CIAP (Calf intestinal alkaline phosphatases) and ligation of DNA with T4 ligase were performed according to the instructions supplied by the manufacturer (Fermentas). Ligation of PCR products to pGEM-T vector was carried out following the instructions of the manufacturer (Promega).

Unless otherwise specified, polymerase chain reaction (PCR) was done using the GeneAmp PCR system 2700 according to [114], under the appropriate conditions in each case. Plasmid DNA isolation was performed with the QIAprep Spin mini prep kit (Qiagen), Nucleospin plasmid kit (Macherey-Nagel) or according to the alkaline method [112]. Purification of DNA and recovery of DNA from agarose gels were performed with the QIAEX II gel extraction kit (Qiagen) or Nucleospin Extract II (Macherey-Nagel).

Agarose gel electrophoresis, fluorescent visualization of DNA with ethidium bromide, spectrophotometric quantitation of DNA or RNA were performed as described by Sambrook [112].

2.3.2.1 Chromosomal DNA isolation from Bacillus amyloliquefaciens FZB42

10 ml LB medium inoculated with a fresh colony was shaken at 37°C and 210 rpm until its O.D.$_{600}$ reached 1. The cell pellet was washed with suspension buffer. Cell walls were destroyed by incubation with 0.7 ml lysis buffer at 37°C for 45 min. 25 µl L-lauroylsarcosine (30%) and 4 µl proteinase K (20mg/ml) were added into the suspension and incubated for 20

min at 70°C in order to degrade cell m embranes and proteins. T o preci pitate proteins and extract pure DNA from the su spension, 700-µl phenol was adde d. After centrifugation for 5 min under 13000 rpm , the aqueous upper layer co ntaining chrom osomal DNA was treated with chloroform/ isoamylalcohol (24:1). The aqueous layer was incubated with 5 µl RNAse at 37°C for 15 m in, then DNA precipitation was performed with 2.5 to 3 volum es of 96% ethanol. The chromosomal DNA was washed with 70% ethanol, and dissolved in 100 µl 1xTE Puffer at room temperature over night.

Buffer	**Compositions**
suspension buffer	10 mM Tris pH 8.0, 150 mM NaCl, 10 mM EDTA
lyses buffer	20 mM Tris pH 8.0, 50 mM NaCl, 10 mM EDTA, 4 mg/ml lysozym
Proteinase K	20 mg/ml in 1xTE buffer
RNase A	100 mg/ml in 1xTE buffer

2.3.2.2 Total RNA isolation

A pre-culture containing 10 ml LB and a fresh colony was shaken at 37°C and 200 rpm for 12 h. It was diluted 1:100 in Landy m edium with 0.25% glucose and shaken at 37°C and 200 rpm until its $O.D_{600}$ reached 0.6. 10 m l of culture was m ixed with 5 m l of Killing buf fer, and centrifuged and 12000 rpm at 4°C fo r 10 m in. After the pellets were washed again with 1 m l of Killing buf fer, they were read y f or total RNA preparation with NucleoSpin RNA L according to the pro tocol of Mach erey-Nagel. In order to concentrate the RNA, one-ten th volume of 1M NaAc (pH 4.5) and 2.5 to 3 volum es of cold 96% ethanol were used for precipitation. The pellets were finally dissolved in 50 µl H_2O containing 0.01% DEPC. A 1.5% agarose denatured gel was used to run RNA samples.

Buffer	**Compositions**
Killing Buffer	20 mM Tris/HCl (pH = 7.5), 5 mM $MgCl_2$, 20mM NaN_3
Loading buffer (2x)	356.25 µl deionized fo rmamide, 106.25 µl form aldehyde, 9375 µl MEN (10X), 20 µg Ethidium bromide
MEN (10x)	200 mM MOPS, 50 mM NaAc, 10 mM EDTA, adjust to pH 7.0 with HAc/NaOH
1.5% agarose denatured gel	0.75 g agarose to be melt in 37.5 m l H_2O , 5 m l MEN a nd 8 m l formaldehyde

2.3.2.3 Southern blot

Southern blotting is a method to detect the presence of DNA by using an agarose gel and hybridization with a labeled probe. In comparison to PCR, this method is not necessary to know the investigated sequence.

a) Preparing labeled probes

Two methods were used to label probes. One was PCR based. Instead of 0.2 mM dTTP, only 0.18 mM dTTP and 0.01 mM digoxigenin-11-dUTP (DIG-dUTP) (Roche, Mannheim) were used. The other way was Klenow fragment 3´-labeling using Ready-To-Go labeling Bead (-dCTP) according to the instructions of the manufacturer (GE Healthcare, Freiburg). The labelled probes were ready for use, after they were purified by a QIAEX II extraction kit (Qiagen) or a Nucleospin Extract II (Macherey-Nagel).

b) Sample preparation and blotting

1-4 µg investigated plasmid or chromosomal DNA was digested overnight with appropriate restriction endonucleases. The DNA was separated on a 0.8% agarose gel in 1x TAE buffer at 60V. First the gel was washed twice in denaturalization buffer for 20 min, then two times es in neutralization buffer for 20 min, finally in 2x SSC buffer for 2 min. DNA was blotted on a Nylon or Nylon^{+} membrane (Macherey-Nagel) under 5 mmHg for 1 hour (Biorad vacuum blotter model 785). DNA was cross-linked on the membrane under UV for 3 min.

c) Hybridization and detection

The membrane was pre-hybridized in 40 ml hybridization buffer for 1 hour at room temperature. For the hybridization, the labeled probe was denatured at 95°C for 3 to 5 min before adding 10 ml of fresh hybridization buffer. The hybridization was performed overnight at a suitable annealing temperature (in this works it was at 55°C).

Before detection, the membrane was washed twice in 2x SSC/0.1% SDS at room temperature for 5 min, twice in 0.5x SSC/0.1% SDS at 55° C for 15 min, once in P1-Dig solution for 1 min, and finally in 100 ml blocking solution at room temperature for 30 min. To detect the labeled probe, the membrane was incubated in 25 ml antibody solution with 3.75 units Anti-Digoxigenin-AP (Roche, Mannheim) for 30 min. The unbound antibodies were subsequently removed after a 15-min washing step. The membrane was neutralized by a 15- min washing step with AP buffer, then incubated in 10 ml AP buffer containing 2.25 mg nitroblue tetrazolium salt (NBT) and 1.75 mg 5-bromo- 4-chloro-3-indolyl phosphate (BCIP). The hybridized DNA with the labeled probe could be visualized on the membrane by incubating in the dark for a while.

Buffer	Compositions
Denaturalization buffer	1.5 M NaCl, 0.5 M NaOH
Neutralization buffer	1.5 M NaCl, 1 M Tris-HCl (pH = 8)
Hybridization buffer	5x SSC, 0.1% (v/v) N-lauroylsarcosine sodium, 0.02% SDS, 1% blocking reagent
SSC (20x)	3 M NaCl, 0.3 M Na-citrate
P1-DIG buffer	100 mM maleic acid, 150 mM NaCl, 1% blocking reagent
Wash buffer	150 mM NaCl, 100 mM maleic acid, 0.3% (v/v) Tween-20
AP buffer	100 mM NaCl, 100 mM Tris-HCl (pH = 9.5), 50 mM $MgCl_2$

2.3.2.4 Reverse transcription PCR

Reverse transcription PCR (RT-PCR) is the preferred technique utilized to detect and quanitfy mRNA. The RevertAid™ M-MuLV Reverse Transcriptase (Fermentas, St. Leon-tot, Germany) and the RiboLock™ RNase Inhibitor (Fermentas, St. Leon-tot, Germany) was used in this work. The RT-PCR mixture contained 1 to 3 µg of total RNA, 1x reaction buffer, 1 µM primer, 1mM dNTPs, 2 units/µl RNase inhibitor and 10 units/µl reverse transcriptase. The total RNA and the primer was firstly incubated at 70°C for 5 min, then the reaction buffer, RNase inhibitor and dNTPs were added. This mixture was incubated at 37°C for 5 min before adding reverse transcriptase. The reverse transcription started with annealing at 42 °C for 1 min, followed by elongation at 70°C for 10 min. To visualize the RT-PCR product in 0.8% agarose gel, an extra standard PCR step using 5-µl cDNA was performed. As negative control, a mixture with 1-µg total RNA was run in parallel.

2.3.3 Genomic methods

2.3.3.1 Shotgun library

The shotgun library was constructed in Göttingen Genomics Laboratory (G2L) in Göttingen. Total genomic DNA was sheared randomly with Nebulizer (Invitrogen), and DNA fragments 1 to 3 kb in size were cloned into pTZ19R or pCR2.2 TOPO to establisha shotgun library. The inserts of the recombinant plasmids were sequenced from both ends using the MegaBACE DNA Sequencing Systems 1000 and 4000 (Amersham-Biosciences) and ABI Prism 377 sequencers (Applied Biosystems) with dye terminator chemistry.

2.3.3.2 Long range PCR

Long range PCR polymerases can amplify fragments longer than that usually attainable by *Taq* DNA polymerase. Mixtures of two polymerases (one with proofreading activity and one without) allow long fragment PCR to be carried out (over 20kb). The length of the primers

was between 30 to 36 bp, and their melting temperature was at 60°C ± 1.5°C. All primers were purified by HPLC. PCR mixtures were consisted with 3 ng/µl chromosomal DNA, 0.5 mM dNTPs, 1x PCR buffer containing 2.5 mM $MgCl_2$ and 0.125 units/µl TaKaRa LA Taq (Shiga, Japan). Denaturation was performed at 94°C for 3 min. In total 30 cycles were performed. The first 10 cycles were done with 20 s of melting at 94°C, 30 s of annealing at 60°C and 20 min of elongation at 68°C. The last 20 cycles were carried out with 20 s of melting at 94°C, 30 s of annealing at 60°C and elongation at 68°C for 20 min with an increase of 20 s/cycle. After the last cycle, the PCR mixture was incubated for further 10 minutes at 68°C to complete the extension of both strands.

2.3.3.3 Multiplex PCR

Multiplex PCR [115] is an essential cost-saving technique for closing a large number of gaps in a whole-genome shotgun sequencing project [116, 117]. The method uses a primer-pooling strategy to minimize the number of laboratory procedures required to sequence the unknown DNA that falls in between contiguous sequences. There are physical limits to how many primers can be used in a single test tube. It was tested that up to 30 primers could be used in one single tube [117].

The primer design is the first critical step in the multiplex PCR. Primers with a melting temperature at 55°C ± 1.5°C were designed for all ends of the contigs from FZB42 genome database. Every primer pool contained randomly 9 primers. Two pools were randomly combined for every PCR. Triple Master Enzyme (Eppendorf, Hamburg, Germany) was used for PCR mixtures that were as follows: 1x tuning buffer with Mg^{2+}, 0.5 mM dNTPs, 1µM per primer, 6 ng/µl chromosomal DNA. Cycling conditions were 96°C for 3 min, 30 cycles of 15 s of melting at 96°C, 30 s of annealing at 55°C, 5 min of elongation at 68° for the first cycle, the elongation time was increased for 10 s per cycle until it reached 9 min.

2.3.3.4 Construction of a fosmid library

a) Chromosomal DNA preparation

The LB culture of *B. amyloliquifaciens* was shaken at 210 until $O.D._{600}$ was approximately 0.8 for chromosomal DNA isolation. MasterPure™ DNA & RNA Purification Kit (Epicentre, Madison, US) was used according the instructions of the manufacturer. DNA pellet was eluted in 35 µl at 50° C pre-warmed H_2O by incubation at 37°C for 30 min or at room temperature overnight.

b) Shearing and End-Repair of the insert DNA

In order to get DNA fragments with approximately 40 kb that are suitable for creating the fosmid library, the genomic DNA was sheared by up and down pipeting for 150-200 times

with a 100-µl tip. A 0.8% agarose gel was used to examine the size of the sheared DNA fragments. End-repair of DNA fragments could be done according the protocol from the CopyControl™ fosmid library production kit (Epicentre, Madison, US), if sufficient fragments with the desired length were there (over 10%). After this treatment, the fragments were ready for blunt-end ligation.

c) Fragment size selection and recovery from the gel slices

To isolate DNA fragments with approximately 40 KB for the ligation, the filled up DNA was separated by a pulse field gel electrophoresis. 1% Low-melt agarose gel (Bio-Rad, Munich) was solved in 0.5 x TBE, which was used also as running buffer. The electrophoresis of the DNA fragments took place in "Two State mode" of CHEF Mapper XA System (Bio-Rad, Munich) with the following parameter: gradient voltage 6 V/cm; included angle 120°; switch time from 1 to 10 s with linear increasing; temperature at 14°C. The running time varied between 22 and 24 h.

The DNA fragments between 38 KB and 42 KB were excised from the pulse field gel. Recovery of the DNA fragments was done using β-Agarase I (New England Biolabs, Frankfurt a.M.). For the optimal digestion of β-agarase I, the gel slices were incubated twice in two volumes of 1x β-Agarase I reaction buffer for 30 min on ice. The gel slices were melted at 65° C for 10 min, subsequently at 42° C for 5 min. 6 µl β-Agarase I was added, after one hour another 3 µl was added. The total incubation time was 90 min. Micron YM-100 column (Millipore GmbH, Schwalbach) was used to completely remove the agarose residues and to concentrate DNA. The column was washed firstly with 500 µl H_2O under centrifugation at 500 xg for 10 min. The digested agarose sample was centrifuged at 16060 xg for 4 min, before loading on the column. The column loading with a sample was centrifuged under 500 xg, until the sample was concentrated in approximately 20-50 µl. The column was washed with 500 µl TE buffer (pH 8) and then with 500 µl H_2O. The wash process was done by centrifugation under 500 xg and the final water volume was approximately 20 µl in the column. The DNA fragments were gathered by putting the top side of the column into a cleaning tube and centrifugation at 1000 xg for 3 min. To avoid the loss of DNA, the column was eluted with another 5-µl H_2O. The entire eluate was completely used for the following ligation.

d) Ligation, Packaging, transfection of fosmids

Since the concentration of the available DNA for ligation was far under guidelines of the Epicentre kit, two different attempts were pipetted. The amount of vector pCC1FOS was used in one approach according to the protocol and in the other one was 1:1 diluted:

10 x Fast-Link Ligations buffer	1 µl	2 µl
10 mM ATP	1 µl	2 µl
pCC1FOS Vector (0,5 µg/µl) 1	µl	1 µl
Concentrated Insert-DNA	6 µl	14 µl
Fast-Link DNA-Ligase	1 µl	1 µl
Bidest. H_2O	-	-
Total volume	10 µl	20 µl

The ligation was accomplished over night at room temperature. 1 µl fresh Fast-Link DNA Ligase was added to the ligation mixture on the next morning and ligation was extended for 45-60 min at room temperature. To inactivate the ligase, the mixture was kept at 70°C for 10 min.

The ligated DNA fragments were packed in the phage using Lambda Packaging Extract (Epicentre, Madison, US), and then infected in the EPI300-T1 $_R$ *E. coli* (Epicentre, Madison, US). Lambda packaging and transfection in *E. coli* were performed according to the instructions of Epicentre.

e) Induction of a high copy fosmid number and fosimid preparation

Induction of the CopyControl fosmids to a high copy number was performed using the standard method of Epicentre. R.E.A.L. Prep Kit (Qiagen; Hilden) was used for fosmid preparation following the protocol of the manufacturer. Quality and concentration of the fosmid were determined by running 0.8% agarose gel and digestion with *Eco*RI, sequentially the fosmid was directly sequenced with pCC1 forward and reverse primer provided by Epicentre.

2.3.4 Biochemical and chemical methods

Unless otherwise specified, the investigated strains were grown in Landy medium [118] for 36 hours at 30°C and 180 rpm. The supernatants were purified by XAD16 for the analysis with bioautography, HPLC and LC-ESI MS. For MALDI-TOF MS, the lyophilized supernatants or colonies were examined.

2.3.4.1 Clean-up the cell culture supernatants

XAD16 is a nonionic macroreticular resin for hydrophobic compounds up to 40,000 MW. XAD16 adsorbs and releases ionic species through hydrophobic and polar interactions;

usually used under isocratic conditions. XAD16 resin was washed with deionized water to remove salts before 10 ml of supernatant was added on the matrix with a 5 beds/min speed. After washing the XAD 16 matrix column with 400 ml deionized water, polyketides and lipopeptides could be eluted with 100% methanol. For the qualitative analysis, 6g of XAD 16 resins and 12 ml 100% methanol were used. In order to analyze the amount of the lipopeptides and the polyketides quantitatively, 12 g resins and 300 ml 100% methanol were used. The methanol samples were dried by using a rotary evaporator. To avoid the degradation of the antibiotics, all the sample were purified at room temperature and neutral pH value. The dried samples were dissolved in 1 ml 100% methanol. These mixtures were ready for the applications in HPTLC, HPLC and LC-ESI MS. The used XAD16 was regenerated by washing with 400 ml methanol (100%), 400 ml acetone (100%) and 400 ml NaOH (1M).

2.3.4.2 High performance thin layer chromatography (HPTLC)

The thin layer chromatography is a separation process, by which substances are isolated according to their different mobility in an organic solvent mixture. The mobility of a substance is dependent on the physical and chemical characteristics of the substance, for instance its size and polarity. The migration of one substance in a certain mobile phase has always the same R_F value which is the ratio of the migration distance (a) of the substance to the simultaneous migration distance (b) of the mobile phase (R_F = a/b).

Samples were concentrated by XAD16. Depending on the sensitivity of the indicator strain, 5-20 µl sample was dropped as point form on a thin layer Silica gel plate (HPTLC aluminium sheet with Silica gel 60 F_{254}, Merck KGaA, Darmstadt, Germany). After plates were run with chloroform/methanol/deionized water (65: 25: 4 v/ v/v) in a sealed chamber, the samples had to be dried completely. The front was marked and the plate was dried horizontally.

2.3.4.3 Bioautography

Bioautography is based on that substances separated on the TLC strips could diffuse into the agar. The active substances that can suppress the growth of a bacterial and fungal indicator showed a clear inhibition zone.

A bacterial indicator was grown under appropriate conditions until $O.D._{600}$ reached about 0.6. Fungus was grown in Erlenmeyer flasks with adequate solid medium, and its spores were gathered through washing with physiological saline solution (0.9% NaCl). The indicator strain was mixed with an appropriate medium containing 1.5% agar. The concentration of the bacterium was $4x10^7$ cfu/ml and the fungus was $4x10^6$ spores/ml. After the medium was solid, the HPTLC strips were put on the plate for 2 hours at the room temperature. After removing the strips, the agar plate was incubated for 8-48 hours at the appropriate temperature.

2.3.4.4 High performance liquid chromatography (HPLC)

HPLC (high performance liquid chromatography) is an analytic method in chemistry, which can not only separate substances, but also identify compounds and their amounts.

According to the stationary phase, there are two methods: normal phase (NP) and reversed phase (RP). A polar stationary phase (e.g. silica gel) is used in the NP-HPLC. A non-polar stationary phase is used in a RP-HPLC, where a polar surface of the silica gel particles is covered with a non-polar layer of an alkane, so the polarity of a RP-HPLC is inverted. RP-HPLC decreases the elution strength by rising polarity in the. As mobile phase, acetone nitrile and tetrahydrofurane (THF) or methanol diluted in water or buffers is usually used. With iso-ionic strength separation, the composition of the mobile phase remains the same during the entire time. With gradient separation, the polarity of the mobile phase is changed during the analysis.

In this work, RP-HPLC was performed with an Agilent apparatus HPLC 1200. ZORBAX Eclipse XDB-C18 was used as the column. The column temperature was kept at 30°C during the experiment. Briefly, a 10-µl sample was injected into the column. The flow rate was 1.5 ml/min solvent A was water containing 0.1% (v/v) formic acid and B was 100% (v/v) acetonitrile containing 0.1% (v/v) formic. The acetonitrile concentration was increased from 10% to 100% within 10 minutes and hold further two minutes at 100%. The column was equilibrated at 5% acetonitrile for 3 minutes after running every sample. A diode-array detector in the range from 190 nm to 550 nm was used for detecting peaks. Polyketides were detected at 280 nm; bacilysin was detected at 230 nm. For the quantitative determination, the area of the peaks was calculated by the auto integrate software of Chemstation from Agilent.

2.3.4.5 HPLC-ESI MS

Liquid chromatography electrospray (HPLC-ESI) mass spectrometry (MS) is able to separate a sample firstly by HPLC. Subsequently, the compounds are vaporized and ionized under a high pressure. HPLC 1 100 Agilent and a Luna 3 µ C18 column were used to perform HPLC. The running conditions were described in 2.3.4.4. Details about ESI MS was described previously [119].

2.3.4.6 MALDI-TOF MS

Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS) is based on ionization of compounds by a laser beam. A matrix is used to protect the compounds from being destroyed by ionization and facilitate vaporization.

To prepare samples for MALDI-TOF MS, supernatants of cultures in Landy medium were lyophilized at 4°C. Those were dissolved in 70% acetonitrile/0.1% trifluoroacetic acid and

mixed with an equal volume of matrix solution. To investigate the compounds that are attached to the outer cell wall, whole cell extraction was performed. Cell material was picked into 100 µl solvent used for dissolving supernatants. Samples were shaken vigorously. After centrifugation, the extracted solutions from supernatants or whole cells mixed with the solvent were spotted on the target and air-dried for the MALDI-TOF MS analysis. Details about MALDI-TOF MS were described in previous publication [120].

2.3.4.7 CAS liquid assay for determining siderophores

The dependency of the cell growth on different iron concentrations is an indirect indicator to know whether bacteria can produce siderophores. This method is not sensitive enough to detect the amount of siderophores. Chrome azurol S (CAS) can form a blue colored complex with iron, which is stabilized by the cationic surfactant hexadecyltrimethylammoniumbromid (HDTMA). Siderophores can deprive iron from the CAS complex, which leads to decolorizing chrome azurol S. The spectrometric measurement of the color alteration is used to determine the siderophore concentration [121].

500µl CAS solution was added to 500µl supernatant from the strains grown in modified Landy medium with 0.1 µM Fe_2O_3. The mixture was incubated for two hours at room temperature in the dark and then O. D. was measured at 630 nm. The ratio of the absorption values from the supernatant/CAS mixture to the one from the Landy medium /CAS mixture reflects the amount of the siderophore in the supernatant [121].

Buffer	**Compositions**
CAS Buffer I (10x)	0.605 % Crome Azurol S, 1 mM $FeCl_3$, 10 mM HCl, 0.729% HDTMA
CAS solution	1x CAS buffer I, 3.024% Pipes, adjust pH to 7.4 with NaOH

3 Result s

3.1 Whole genome B. amyloliquefaciens FZB42 sequencing project

Bacillus amyloliquefaciens FZB42, a Gram-positive rhizospher e bacterium, is applied world-wide in farming and gardening, and shows a significant effect on enhancing plant growth. In order to identify the genes involved in this pr ocess, the genome of FZB42 was sequenced and compared to the genome sequence of *Bacillus subtilis* 168, a well-known laboratory m odel strain that is not endued with plant growth prom otion capabilities [7]. In this chapter, I will present the results that I got in the process of genome sequencing.

3.1.1 Shotgun sequencing

The shotgun m ethod m akes it possible to get th e m ain part of the inform ation from the genome sequence in a very short time. Göttingen Genome Laboratorium (G2L) performed the Shotgun sequencing and assembling of the reads. 39,175 reads were generated by the shotgun approach. Af ter the f irst autom atically ass embling by Phrap (http://www.phrap.org/), 814 contigs were generated in the genome database, which covered 3,836 kb.

3.1.2 Editing contigs

Contig editing was supported by G AP4 (http://staden.sourceforge.net/). The m ajor tasks f or editing contigs were solving problems of m isassembling and si ngle nucleotide base error in contigs.

Figure 9. An example of misassembly in the gap4 Contig Editor window: The disagreement of base between the consensus and individual reads are shown in green. The confidence of a base in a consensus is described by the grey scale value. The confidence reduces from light to dark.

Misassembling is due to repeats in the genomic DNA, unclonable sequences, chimeric clones, etc. (**Figure 9**). All of the misassembled sequences were identified by checking the contigs in the Contig Editor . Thes e read s were m anually disassem bled and th en reassem bled into a correct contig.

Single nucleotide base error existed very frequently in contigs. These tiny widely spread errors might result in frameshift in an open reading frame and affect protein prediction. The point mutation, bad quality of reads and low coverage of reads were the reasons for this problem. To avoid the spontaneous point mutation in the genome sequence, the chromosomal DNA was isolated always from a bacterial stock kept in -80 °C. To fish mistakes of bases in contigs, the contig containing the whole genome sequence was checked manually from beginning to end; regions with insufficient sequence quality (base confidence lower than 65%) were identified and improved through re-sequencing.

3.1.3 Gap closure

The most challenging step in the sequencing project was gap closure, especially a sequence with a big span of repeats like rRNA operon. In order to close these gaps, several strategies were used in this work.

3.1.3.1 Gap closure by Gap4

Gap4 (http://staden.sourceforge.net/) is not only useful for editing sequences but also one of the most convenient tools as a first step to conquer the problem of gap closure.

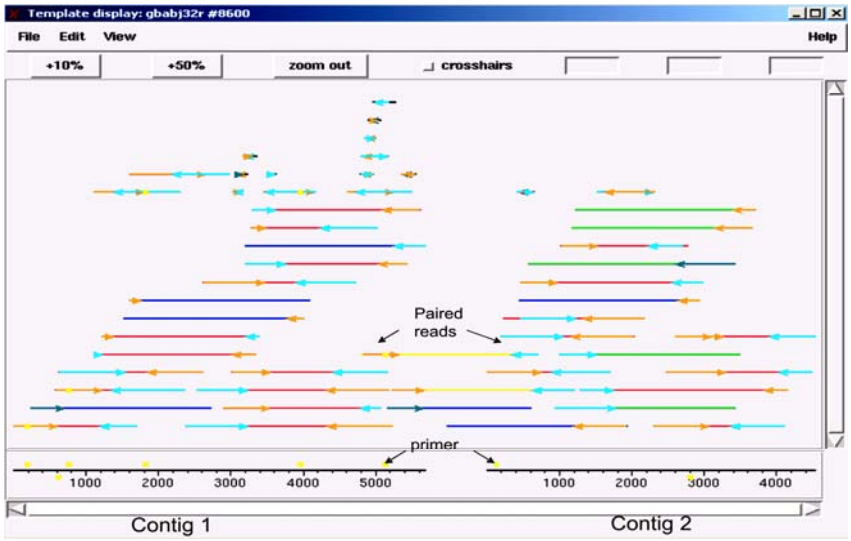


Figure 10. Gap closure by using Gap4 software: Two contigs were shown in Contig Editor. Arrows represent reads (blue for reverse reads and orange for forward). Red line is for the forward and the reverse read from one plasmid connected in a same contig, yellow for the two relative reads belonging to two different contigs with an appropriate distance (< 3.5 kb), green for the one with hypothetical distance bigger than 4 kb, blue line for missing complementary read.

Gap4 could find two neighbor contigs, if a forward read and its reverse read came from the same plasmid and were located in two different contigs. It was important that the expected distance between these two reads should be smaller than 3.5 kb, because an insertion of a chromosomal DNA fragment in a plasmid from the shotgun library was about 3 kb. Primers

could be designed on the related plasmid and PCR walking was performed. If the assumed distance of these two reads was bigger than 4 kb, it indicated that at least one of the reads was in a wrong contig, namely misassembling (**Figure 10**) which had to be corrected by reassembling.

3.1.3.2 Gap closure by comparison with B. subtilis *168*

The phylogenetic tree analysis (**Figure 11**) indicated that *Bacillus amyloliquefaciens* FZB42 is the closest one to *Bacillus subtilis* 168 and followed by *Bacillus licheniformis* ATCC14580. Therefore, it was presumed that the organization of the genetic information from FZB42 was similar to BS168. All the proteins from BS 168 were compared by tblastn against the nucleotide database of FZB42 containing all contigs. If one or two genes in the neighborhood of the *B. subtilis* 168 genome found their similar counterpart locating in the different ends of two contigs, at least two pairs of the primers were designed on the two ends of the related contigs. If the PCR-product existed, the primers were used for further sequencing.

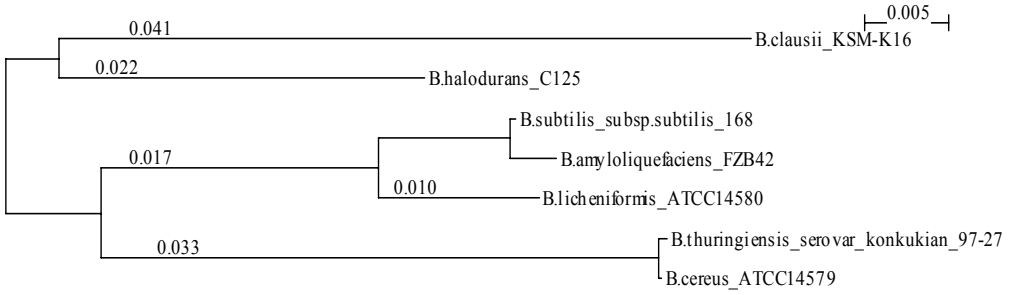


Figure 11. Phylogenetic tree of some members of the genus Bacillus based on 16s rRNA sequence analysis

3.1.3.3 Primer walking

If an interesting sequence obtained from PCR was bigger than 1.5 kb, which is common in the case of long range PCR, it was too long to be sequenced at one time by using primers from both ends of a fragment. In this case, primer walking was used, where new primers were designed by using the information from the prior end of sequences. This process was repeated, until the information from the whole DNA fragment was obtained.

After the first round of gap closure using the methods described above, the total number of contigs reduced to 228, among which 71 were bigger than 3 kb.

3.1.3.4 Gap closure by software MUMmer

Comparison with the closest related species *Bacillus subtilis* 168 accelerated the progress of gap closure. To exploit this method for gap closure, the genomes of other closely related species (*B. licheniformis* ATCC14580, *B. cereus* ATCC14579, *B. clausii* KSM-K16, *B.*

halodurans C125, *B. thuringiensis serovar konkukian* 97-27) were compared with the one of FZB42.

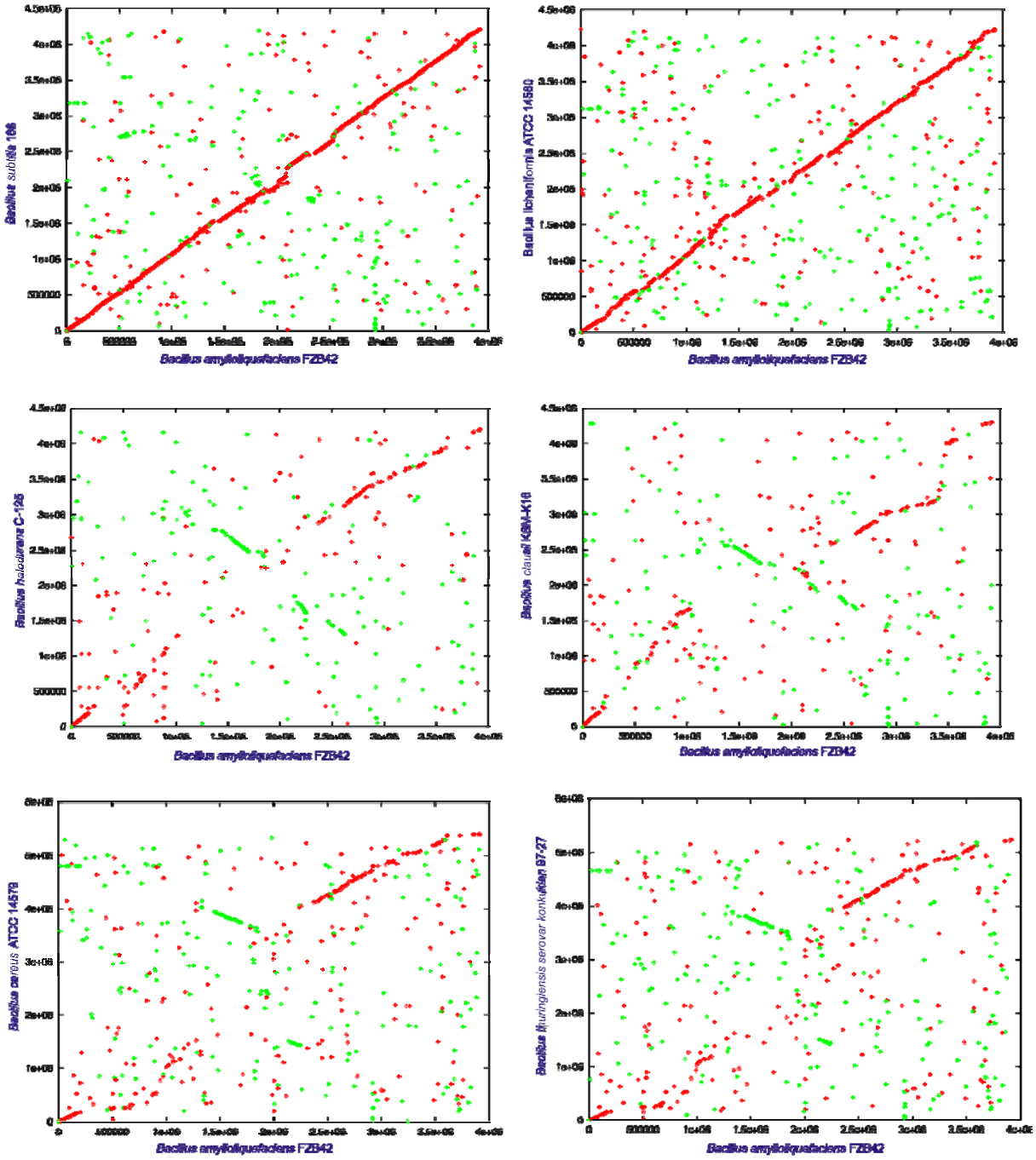


Figure 12. Alignment of the whole genome se quences by using MUMmer: the X -axis stands for the whole genome sequence of Bacillus am yloliquefaciens F ZB42 and the Y -axis for the compared genome seque nce. A dot is a sequence th at is matched in both of the genomes. Th e sequences can be matched on the forward (red dots) or the reverse (green dots) strand.

It would have been very tim e consuming, if the comparison had been done m anually. For this purpose, the whole genom e alignment program MUM mer (http://mummer.sourceforge.net/) was applied, which can find all m aximal, unique, matching sequences (MUMs). The longest

possible set of matches that occur in the same order (**Figure 12**) was extracted. This information was used for contigs ordering that was confirmed subsequently by primer walking. After this round of gap closure, 179 contigs remained in the Gap4 database, which contained 48 contigs longer than 3 kb and 2 contigs longer than 400 kb.

This comparison confirmed the conservation of organization of genetic information in very closely related strains (like *B. subtilis* and *B. licheniformis*) and indicated "hot-spot" rearrangement loci, where gene transfer could have occurred. These "hot-spot" rearrangement loci exist mainly at the end of the replication fork (**Figure 12**).

3.1.3.5 Gap closure by using fosmid library

To conquer the problem with a large scale of gaps between the contigs, it was necessary to establish a clone library that contains fragments about 30 kb in size. At first, a cosmid library was tried, but no clone was obtained after several attempts, probably due to instability of the cosmids. In contrast to cosmids, fosmids have a lower copy number and higher stability. Using fosmids, a genome library containing 96 clones with fragments larger than 25 kb was established. 249 reads were obtained directly from the fosmid library; one of the fosmids was sub-cloned to get a shotgun library for further sequencing.

3.1.3.6 Gap closure by multiplex PCR and long range PCR

After exploiting the fosmid library, the number of the contigs larger than 1000 bp was reduced to 31. For the remaining contigs, there was no evidence for contig ordering. All these contigs had to be tested randomly by PCR to assess contig order. In this case, multiplex PCR was a powerful method, which enabled simultaneous amplification of many interesting targets in one reaction by using more than one set of primers [122].

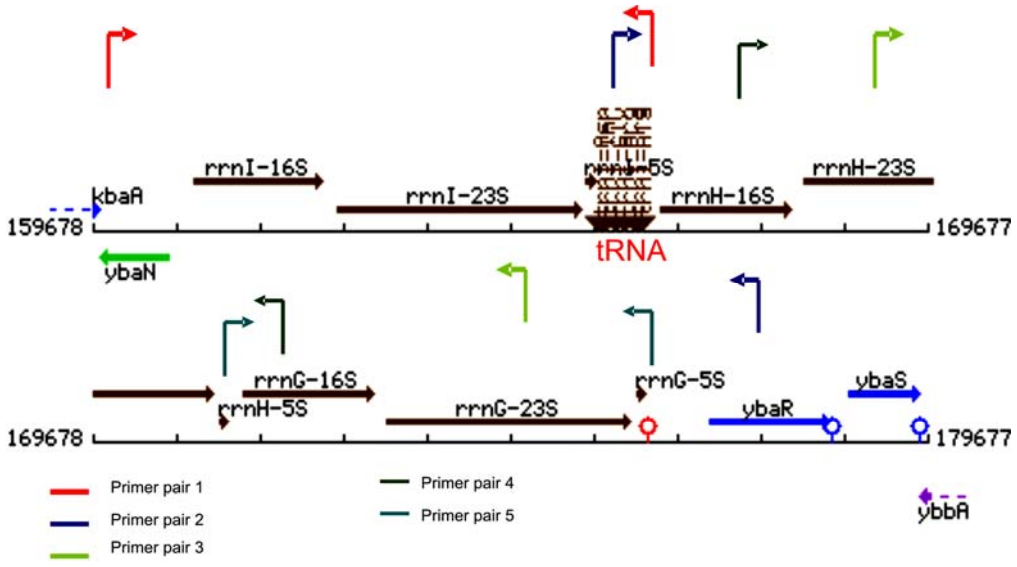


Figure 13. Strategy for gap closure containing at least two rRNA operons

After using multiplex PCR there were still three gaps remaining among the contigs that were larger than 400 kb, from which four ends contained rRNA operons. Two of the gaps were closed by long range PCR using a unique primer at the end of one contig and a non-unique primer at the end of another contig containing rRNA operon in the end. The last gap in the genome sequence was the most difficult one, because both ends contained rRNA operon, one of the contigs contained two rRNA operons and the total number of rRNA operons in this gap was unknown. The sequence around this gap was found conserved very well in comparison with BS168. It could be hypothesized, that the number of rRNA operons in FZB42 was the same as in *Bacillus subtilis* 168. To confirm this assumption, three primer pairs were designed according to the sequence of *B. subtilis 168*. Primer pairs 1 and 2 were used to amplify two specific PCR-products on FZB42 chromosomal DNA, whose size were similar to that of *B. subtilis* 168 (Figure 13). It ruled out that three rRNA operons exist in this region. Primer pairs 3, 4 and 5 were used to check the direction of the operons (**Figure 13**). PCR and sequencing with these five primers confirmed that the number of rRNA operons and their organization are identical to that of *B. subtilis* 168.

3.1.4 Sequence analysis of the genome of *B. amyloliquefaciens* FZB42

The process of identifying location and function of open reading frames (ORFs) is called annotation. ORFs of FZB42 were identified firstly by direct manual comparison of the DNA sequence against the genome database of *B. subtilis* 168 (genolist.pasteur.fr/SubtiList). This result was then combined with the automatic ORF prediction done by G2L using YACOP program which contains the ORF-finding programs Glimmer, Critica and Z-curve [123]. 3984 ORFs were identified in FZB42 after automatic searching. All predictions were verified and modified manually by confirming the presence of ribosomal binding site, codon usage analysis and sequence homology analysis by the software GeneSOAP [124]. The sequence homology analysis was done by comparing coding sequences against public databases, like SwissProt, GenBank, ProDom, COG, Prosite and KEGG.

During the annotation, frameshifts were detected mainly due to sequencing mistakes and point mutations. Sequencing mistakes occurred because of bad quality of sequences or contigs, while point mutations happen spontaneously in nature, which often lead to gene silence. To rule out the frameshifts that did not be caused by artificial sequencing mistakes, all frameshift regions were examined again by sequencing. After completing the annotation, the final number of coding sequences (CDS) is 3693. Three types of protein annotation were assigned: 1) 2340 (63.3%) were proteins with putative functions; 2) 1141 (30.9%) were conserved hypothetical proteins whose homologues were found in other genomes, but their functions

were still unknown; 3) 214 (5.8%) were hypothetic al proteins that were unique for FZB42. The genom ic infor mation of *B. amyloliqu efaciens* FZB42 is availab le in the NCBI databas e under the accession number gi|154350369.

The m anual annotation provides a great am ount of genetic, biochemical and physiological data, e.g. metabolism pathway , enzym e activit y and dom ain or ganization, which gives the hints about the functions of the genes. However , their exact functions should be investigated by further experiments.

3.2 Genomic analysis

Genome analysis of FZB42 has been extensivel y described in several previous publications [Lit. 1, Lit. 6, Lit. 7 and Lit. 8] Here I will present a short summary and stress some additional aspects of my genome analysis.

Table 1. Gen omic features of B. am yloliquefaciens in com parison with the genomes of other Bacillus strains:

	B. amyloli-quefaciens FZB42	*B. sub tilis* 168	*B. l icheni-formis* ATCC 14580	*B. cla usii* KSM-K16	*B. halodurans* C-125	*B. cer eus* ATCC 14579	*B. thuringiensis serovar konku-kian 97-27*
Genome size (bp)	3,918,589	4, 214,630	4, 222,645	4, 303,871	4,202,352	5, 426,909	5,237,628
G+C content (mol%)	46.4	43 .5	46 .2	44	43.7	35 .4	35.4
Protein coding sequences	3,693	4, 106	4, 196	4, 096	4,066	5, 366	5,263
Average CDS size (bp)	933	89 5	87 7	90 0	879	83 5	826
Percent of coding region	88.0	87 .2	87 .2	85 .0	85.0	84 .0	83
Ribosomal RNA operons	10	10	7	7	8	13	13
Number of tRNAs	89	86	72	74	78	108	105
Phage-associated genes	44	268	71	18	42	124	11
Transposase genes of IS elements	9	0	10	22	93	10	60

Table 2. Substitution of prophages in the genome of FZB42 in comparison with B. subtilis 168 [Lit. 6]:

From (BS[1])	To (BS[1])	Size (BS[1])	RBAM DR	[2]	phage tRN	A	remark
200263 *glmS*	228161 *csgA*	28898	RBAM183 RBAM206	+	Prophage 1	-	RR/HK, detoxification (export antimicrobial peptides), KDPG reductase, RR/HK, ABC transporter; substitution of prophage 1(deletion 1: 24492 bp)
502777 *ydcK*	521213 *cspC*	31118	RBAM460 RBAM482	+	Prophage 2	Leu	ester cyclase, transporter, MarR regulator (2), NAD dependent dehydrogenase, NADP quinone reductase, TetR, regulator, oxido reductase (3), fatty acid desaturase, substitution of prophage 2 (deletion 2: 19208 bp)
524801 *yrkE*	560664 *ydeR*	35863	RBAM485 RBAM525	+	Prophage 3	-	rhodanase, aspartate racemase, PadR regulator, Mannose-6-phosphate isomerase, dihydrodipicolinate synthase, Fe-S-cluster redox enzyme, pyrimidine kinase, sugar phosphatase, arsenite efflux pump, Na+H+ antiporter, GntR regulator, PaiB regulator, methyl transferase, RR/HK, substitution of prophage 3 (deletion 3: 30805 bp)
1164638 *yjcJ*	1193383 *yjfB*	28745	RBAM1106 RBAM1135	+	Prophage 4	Val	sugar transporter, galR, ganA, galT, galE, galK, ptsIIBC, ptsIIA, P-ß-gal, deoR regulator, phosphohexomutase, substitution of prophage 4 (deletion 4: 28412 bp)
1317799 *xlyB*	1347491 *xlyA*	28786	RBAM1158 RBAM1177	+	PBSX	-	rearrangement of prophage related genes yomR and yomQ (deletion 5: 19751 bp)
1793935 *glnA*	1796733 *xynP*	2798	RBAM1630 RBAM1636	+	phi105	-	phage integrase, gluconolactonase
2050542 *pps*	1916851 *lexA*	133691	RBAM1641 RBAM1669	+	Prophage 5	-	homing endonuclease, nuclease inhibitor, dehydrogenase, protein acetyltransferase, chitin binding, transporter, regulator, substitution of prophage 5 (deletion 6: 24681 bp)
2078498 *yodU*	2095572 *yokA*	134777	RBAM1863 RBAM1881	+	SPßC2	-	Replication terminator yodN, phage proteins but lacking major SPßc2, RR/HK, site specific recombinase YokA, , partial substitution of SPßC2 (deletion 7:134777 bp)
2652600 *spoIVCB*	2700977 *spoIIIC*	48377		+	skin element	-	insertion of putative polysaccharide deacetylase RBAM02462 and unknown gene RBAM02464 (deletion 8:48377 bp)
3767647 *yxdJ*	3772893 *fba*	5246	RBAM3511 RBAM3517	-	Imo828	-	Detoxification: mersacidin export and regulation, mrsK, mrsR, mrsF, mrsG, mrsE

1. BS: *Bacillus subtilis 168*
2. DR: direct repeats

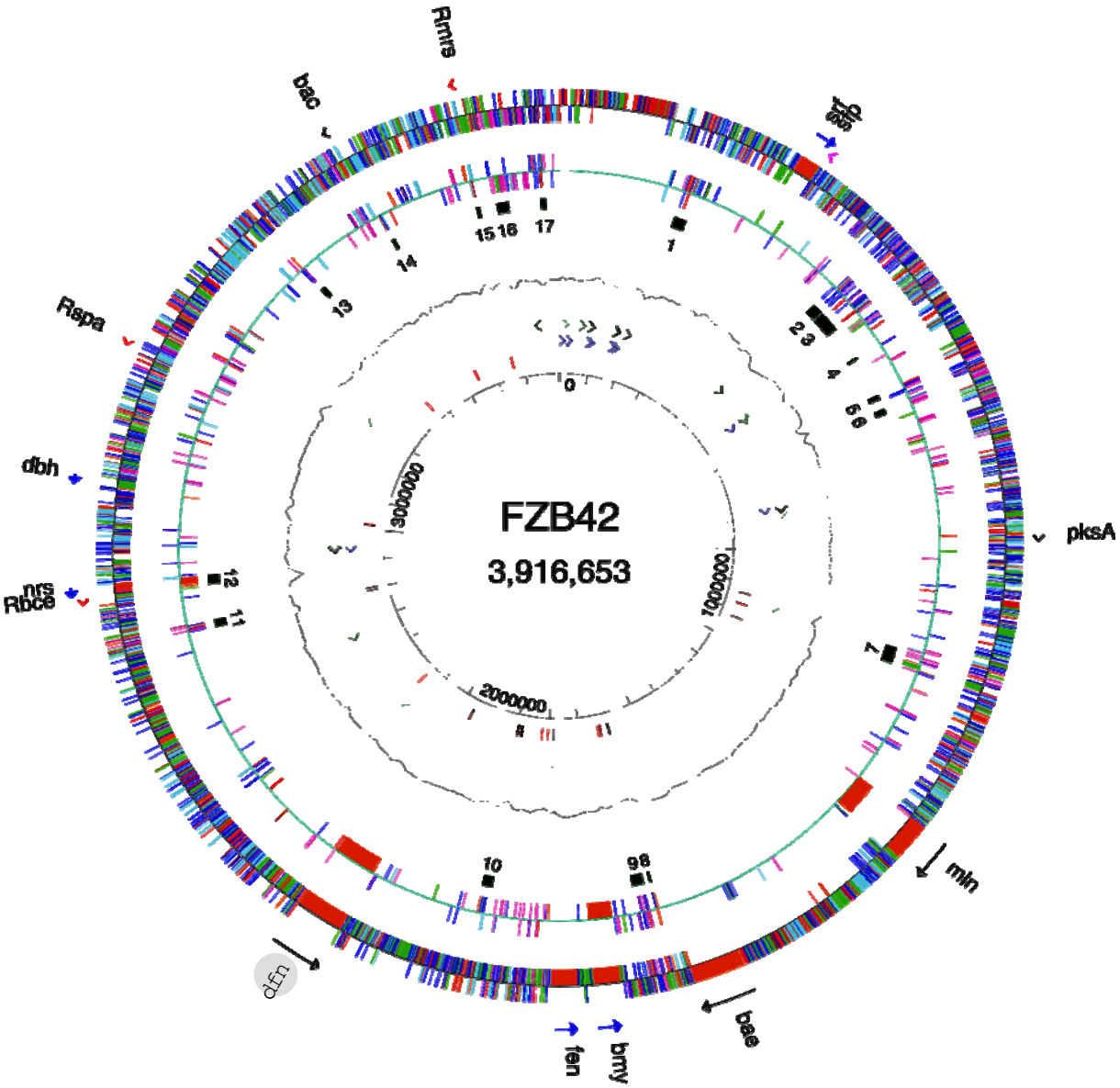


Figure 14. The whole g enome of Bacillus amylo liquefaciens F ZB42: Genes and gen e clusters involved in synthesis and export (detoxification) of secondary metabolites l ie in the most outside of the cir cle. 1st cir cle: all genes in color code ac cording to their functions: cell enve lope and cellular pr ocesses, gr een; information pathways, orange; intermediary met abolism, pink; other functions, r ed; unknown, black. 2nd circle: Gene s not conserved in B. subtilis including four giant gene clusters invol ved in synthesis of secondary metabolites (orange); 3rd ci rcle: the numbered 17 DNA islands (green), 4th circle: GC-content profile, 5th circle: rRNAs (green), 6th circle: tRNAs (cyan), 7th circle:prophages (black), transposons, and IS elements (red), 8th circle: scale (bp) [Lit. 6].

3.2.1 General features of FZB42 genome

FZB42 possesses an AT-rich genome with 3,916,653 bp. General properties of its sequence in comparison to other genom e sequences of related *Bacilli* are shown in Table 1. Addition al genomic information of other *Bacilli* is available from the NCBI genomics projects database (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi).

The majority of the FZB42 CDS are conserved in the closely related *B. subtilis* (3,181) and *B. licheniformis* (2,857), and most of them are arranged in a collinear manner in all three strains (Table 1, **Figure 12** and **Figure 14**). These three strains share 2494 conserved genes totally. Genes involved in signal transduction, sigma factors, competence, transporter systems, secretion systems and extracellular proteins were summarized in [Lit. 6].

3.2.1.1 Absence of prophage related genes

Unlike *Bacillus subtilis* 168, the total amount of A (26.7%) and T (26.8%), G (23.3%) and C (23.2%) is similar in FZB42. In comparison to BS168, the genome of FZB42 is about 300 kb smaller, mainly due to absence of A+T-rich prophage related genes. In total, BS168 contains four known prophages (PBSX, SPβ, skin element and phi105) and six prophage-like elements [7]. In FZB42 330503 bp were absent, compared to BS168. The prophages, *skin* element (48 kb) and SPβC2 (134 kb), were deleted almost completely in the genome of FZB42 (Table 2). Notably, loss of production of the anti-listerial bacteriocin subtilosin in FZB42 is due to deletion of the corresponding genes within the SPßC2 region; however, FZB42 contains the genes for four extra non-ribosomally synthesized compounds (see 3.2.1.2).

3.2.1.2 Numerous secondary metabolites produced by B. amyloliquefaciens *FZB42*

A remarkable feature of FZB42 is its capacity to produce a high diversity of antibiotics. Nine giant gene clusters, which are responsible for non-ribosomally synthesized secondary metabolites, are present in the *B. amyloliquefaciens* FZB42 genome: three lipopeptides (surfactin, fengycin and bacillomycin D), three polyketides (*pks1*, *pks2* and *pks3*), one dipeptide (bacilysin), one siderophore (bacillibactin) and two unknown compounds (**Figure 14**). One of the unknown compounds is a PKS/NRPS hybrid; the other is Sfp-independent and was detected by antibiotic resistance gene analysis [125].

3.2.2 Genomic analysis of gene clusters involved in production of secondary metabolites by *B. amyloliquefaciens* FZB42

3.2.2.1 Lantibiotics and other ribosomally synthesized antibiotics produced by B. amyloliquefaciens *FZB42*

Despite that *Bacilli* produce a series of lantibiotics (for details see 1.3.1), no structure genes of known lantibiotics are present in the genome of FZB42 (Table 3). However, some of the transporters, regulators and immunity genes for lantibiotics were identified (Table 3). For example, the signal transduction genes *mrsK2R2* involved in regulation of biosynthesis of mersacidin and their immunity genes *mrsEFG* are present (Table 3). A similar gene cluster containing the signal transduction genes *spaKR* and the immunity genes *spaEFG* of subtilin exists. In case of subtilosin A, only a putative transporter, AlbG, was found, whilst the

immunity genes *albBCD* are not present. The immunity and regulation genes of sublancin were not characterized until now. However, the immunity genes of sublancin are thought to be localized in the region of the sp β2 prophage [126, 127] which is completely missing in the genome of FZB42 indicating that it does not contain specific immunity genes against sublancin (Table 3). No signal transduction genes, transporters and immunity genes directed against ericin are present in FZB42 (Table 3).

A ribosomally synthesized unknown compound that shows antibiotic activity against *B. subtilis* HB0042 (*sigW*) was detected by screening of *sigW*-dependent resistance genes [125]. *ydbST* and *fosB* were identified as the resistance genes for the novel antibiotic in *B. subtilis*. Homologues of the *ydbST* genes have been identified in a plasmid of *Staphylococcus aureus*, they are involved in production of, and immunity to, the bacteriocin aureocin A53 [125]. *fosB* is a known immunity gene for fosfomycin. Analyzing the whole genome sequence of FZB42, no genes were detected, which could be involved in the biosynthesis of aureocin-like or fosfomycin-like antibiotics.

Table 3. Presence of lantibiotics and related genes in B. amyloliquefaciens FZB42:

Name of the lantibiotics	Genes involved in biosynthesis of the lantibiotics presented in their producers						Lantibiotics related genes presented in FZB42
	Genes	Structure gene	Post-transcriptional modification	Transporter	regulatior	immunity	
mersacidin	*mrsK2R2 EFGADMT*	*mrsA*	*mrsMD*	*mrsT*	*mrsK2R2*	*mrsFEG*	*mrsk2R2 mrsEGF*
subitilosin A	*sboAX albABCDEFG*	*sboA/ spoX?*	*albA*	*albFG*		*albBCD*	*albG*
sublancin	*sunAT/bdbA yolJ/dbdB*	*sunA*	*bdbAB*	*sunT*			-
subtilin	*spaBTCS FEGRK*	*spaS*	*spaBC*	*spaT*	*spaRK*	*spaFEG*	*spaFEG spaRK*
ericin	*eriBTCAS FEGRK*	*eriAS*	*eriBC*	*eriT*	*eriRK*	*eriFEG*	-

The data from the transposon library of CH5 (*sfp*$^-$ and *yczE*$^-$) showed that a unique gene cluster in FZB42 is involved in production of this antibiotic (Zhiyuan Wang and Anto Budiharjo, unpublished). This gene cluster is inserted between *guaC* and *yutM* counter-clockwise to the DNA replication (Figure 15).

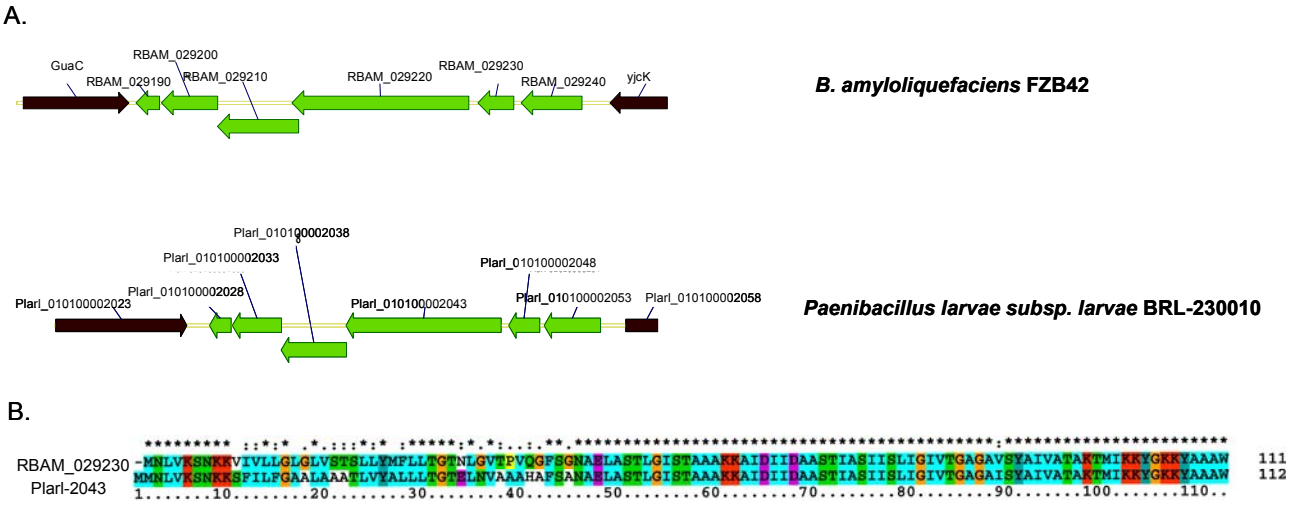


Figure 15. The gene cluster related to a novel antibiotic in FZB42 is very similar to a gene cluster involved in the genome of Paenibacillus larvae subsp. larvae BRL-2300: A. Gene organization in FZB42 and BRL-2300. The green marked genes are related to production of this novel antibiotic, the brown-labelled genes are in the neighbourhood to this gene cluster. B. Alignment of RBAM_029230 and Plarl_2043 shows that this gene is highly conserved in both strains.

This operon contains six genes and its organization is similar to that of a gene cluster present in *Paenibacillus larvae subsp. larvae* BRL-230010 [128], an endospore-forming bacterium causing American foulbrood disease in honeybee (**Figure 15**A). The similarity of these genes on the amino acid level is ranging from 46% (RBAM_029240 to Plarl_2053) to 80% (RBAM_029230 to Plarl_2043) (Table 4). Notably, the C-terminii of RBAM_029230 and Plarl_2043 are almost identical (**Figure 15**B).

Table 4 Features of the proteins involved in production of this novel antibiotic:

	length (bp)	MW Dalton	Conserved domains sign	alP[1]	secretomeP[2]	No. of TMH[3]	similarity to Plarl[4]	protein in Plarl[4]
RBAM_029240	570	21138	n. d.[5]	30/31[6]	Yes	5	46%	2053
RBAM_029230	336	11350	n. d.[5]	39/40	Yes	2	80%	2048
RBAM_029220	1653	62891	n. d.[5]	n. d.[5]	Yes	12	51%	2043
RBAM_029210	702	25971	ABC-transporter	n. d.[5]	no	0	72%	2038
RBAM_029200	447	16290	integral membrane protein	n. d.[5]	no	3	56%	2033
RBAM_029190	219	8311	n. d.[5]	n. d.[5]	Yes	3	55%	2028

1. signalP predicts whether there is signal peptide in the protein sequence and its possible cutting position.
2. secretomP predicts whether the protein can be secreted without signal peptide.
3. TMH: transmembrane helix
4. Plarl: *Paenibacillus larvae subsp. larvae* BRL-230010
5. n. d.: not detected
6. Probable cleavage site in the position of the amino acids

To predict the structural gene of this unknown antibiotic, conserved domains were analyzed in this unique gene cluster . The results s howed that RBAM_029200 contained an integral membrane protein domain with unknown f unction and R BAM_29210 contained an ABC-transporter dom ain belonging to the BcrA superfam ily (Table 4). This im plied that RBAM_029200 and RBAM_29210 could not be the structural gene whose product is secreted out of the cell or anchor to the outer of the cell membrane. The rest proteins in this cluster did not contain any known conserved do mains. Prediction software for signal peptides (SignalP), extracellular proteins wit hout signal peptide (Secretom P) and transm embrane helix (TMHMM) (CBS Pre diction Servers, Denmark) were used to an alyze th ese proteins. RBAM_029230 and RBAM_029240 were predicted to contain a signal peptide or a signal anchor (Table 4). The p robability th at RBAM_02940 had a signal anch or was 0.43 3 and a signal peptide 0.559, while the probability that RBAM_029230 contained a signal peptide was one, which m eant that RBAM_029230 contained a signal peptide with a possibility of 100% (Table 4). The results from secretomP analysis im plied that RBAM_29220 and RBAM_29190 belong to the secretom e protein, despite that signalP prediction gave no evidence for presence of signal p eptides and signal anchors in these proteins (Table 4). However, RBAM_29220 and RBAM_29190 cannot be a secreted protein, because the m ain parts of these proteins are transmembrane helices (Table 4).

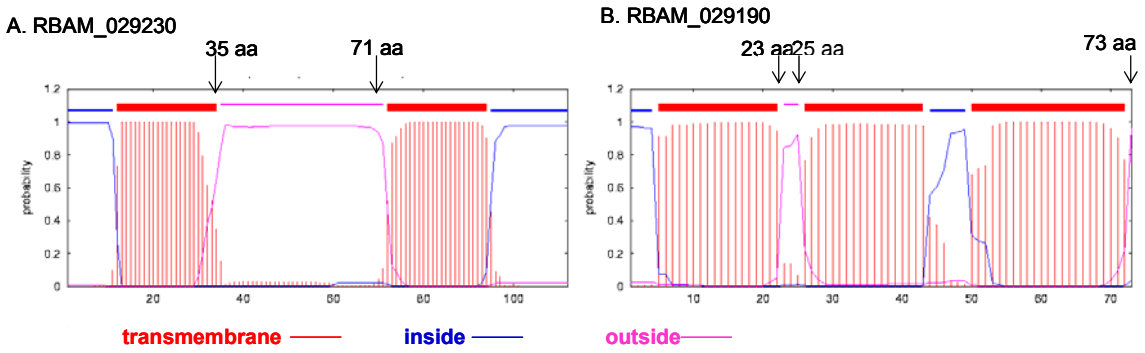


Figure 16. Transmembrane helix analysis of RBAM_029230 and RBAM_029190

The results from signalP and secretom P showed that RBAM_029230 and RBAM_29190 could be candidates for struct ure genes. RBAM_29190 contained two tiny extracellular parts (**Figure 16** B), which implied that RBAM_29190 was un likely a structural gene for this unknown antibiotic. In com parison to it, transm embrane helix (TMH) analysis illustrated that RBAM_029230 contains two TMHs, between which there are abou t 36 am ino acids outside of the cytoplasm (**Figure 16** A). It seem s that RBAM_029230 could be the structural gene for this antibiotic and would be cut at the positi on of the 36th aa of the N-term inus and the 41th

aa of the C-terminus to form a hypothetical mature protein, before it is released from cells (**Figure 16**A). Features of the mature protein were investigated by Protein Calculator v3.3 (the Scripps Research Institute, La Jolla, USA): it would have a molecular weight of 3455.8601 Da and pI 5.69. Cysteine, the characteristic amino acid for lantibiotics, is not present in the mature protein. Because of lacking tryptophan, tyrosine and disulfide bond, it has theoretically no absorbance at wavelength 278 nm to 282 nm.

HPLC, HPLC-ESI MS and MALDI-TOF MS analysis yielded no peaks corresponding to the predicted size of this unknown antibiotic in the supernatant and/or on the surface of the cytoplasm membrane.

3.2.2.2 Nonribosomally synthesized lipopeptides produced by B. amyloliquefaciens *FZB42*

Three lipopeptide synthetases cover 101,077 bp of the genome of *B. amyloliquefaciens* FZB42. The *bmy* gene cluster lies 1.8 Mbp and the *fen* gene cluster 1.9 Mbp distant to the DNA replication start point in counterclockwise orientation; the *srf* gene cluster locates 0.3 Mbp clockwise to the replication start point.

The 26.1-kb *srf* operon present in the *B. amyloliquefaciens* FZB42 genome is organized similarly as in *B. subtilis* 168. The corresponding genes of these two strains exhibit identity between 72% (*srfAA*) and 83% (*srfAC*) at the amino acid level. Like *B. subtilis* 168, the genes present at the left flanking region of the operon are *hxlBAR*, while the genes at the right flanking region of the *srf* gene cluster differ slightly to the ones from *B. subtilis* 168. The two genes *ycxAB* with unkown functions in BS168 are substituted by two ORFs (RBAM_003700 and RBAM_003710) with unknown function in FZB42. RBAM_003700 contains a conserved domain, belonging to aminotransferase class I and II, and exists in other *Bacilli*, while RBAM_003710 did not show any similarity to known genes in the database.

The *fen* gene cluster is about 37.6 kb and contains five genes (*fenABCDE*). It locates in *B. amyloliquefaciens* FZB42 at the same locus as the *pps* operon in *B. subtilis* 168. The homology of *fen*ABCDE to *pps*ABCDE is between 61% and 66% and this region is conserved well in both *Bacillus* species.

The 37.2-kb *bmy* gene cluster (*bmyABCD*) does not exist in BS 168, and is inserted between *xynD* and *yngA* with a rearrangement of the biotin synthesis gene cluster. Notably, the genes in the neighborhood of the *bmy* operon are the same as in the iturin A gene cluster in *B. subtilis* RB14 [64] and the *bmyL* gene cluster in *B. subtilis* A1/3 [129]. BmyA shows a high level of homology to ItuA (98%) of *B. subtilis* RB14, BmyB to ItuB (81%) of *B. subtilis*

RB14 and BmyC to MycC (74%) of *B. subtilis* ATCC6633. A more detailed description of the lipopeptide gene clusters in FZB42 has been given elsewhere [Lit. 8 and Lit. 1].

3.2.2.3 Polyketides produced by B. amyloliquefaciens *FZB42*

Three PKS gene clusters covering 196,340 bp are localized at the positions around 1.4 Mbp (*pks2*), 1.7 Mbp (*pks1*) and 2.3 Mbp (*pks3*) in clockwise direction to the DNA replication origin of the *B. amyloliquefaciens* FZB42 genome.

Table 5. List of pks1 ORFs and their homologous proteins:

Protein/Gene	Size (Da/bp)	Similar to strain	Similarity [%]	Homolog protein	Domain
BaeB/*baeB*	25875/678	*BS* 168	64	PksB	GloB
BaeC/*baeC*	32317/870	*BS* 168	72	PksC	AT1
BaeD/*baeD*	36425/975	*BS* 168	52	PksD	AT2
BaeE/*baeE*	82426/2241	*BS* 168	69	PksE	AT3, oxidoreductase
AcpK/*acpK*	9306/249	*BS* 168	69	AcpK	ACP
BaeG/*baeG*	46539/1263	*BS* 168	83	PksG	ACP
BaeH/*baeH*	29202/774	*BS* 168	65	PksH	ECH (enoyl-CoA hydratase)
BaeI/*baeI*	27788/750	*BS* 168	75	PksI	ECH (enoyl-CoA hydratase)
BaeJ/*baeJ*	551392/14949	*BS* 168	62	PksJ	[AL0, ACP0], [C1,A1,PCP1], [KS2,DH2, KR2, ACP2,ACP2], [KS3, KR3,ACP3], [$KS4^0$
BaeL/*baeL*	495437/13428	*BS* 168	63	PksL	DH4,ACP4], [KS5,DH5,KR5,ACP5], [KS6, ACP6-1, ACP6-2], [KS7,KR7, ACP7], [$KS8^0$
BaeM/*baeM*	389407/10536	*BS* 168	62	PksM	DH8,ACP8], [KS9,DH9,KR9, MT9, ACP9], [KS10,KR10,ACP10]
BaeN/*baeN*	599995/16302	*BS* 168	63	PksN	[C11,A11,PCP11], [KS12,DH12, KR12, ACP12], [KS13, DH13,KR13, ACP13], [KS14,DH14,KR14
BaeR/*baeR*	276050/7449	*BS* 168	57	PksR	MT14,ACP14], [KS15, DH15, ACP15], [KS16, ACP16,TE]
BaeS/*baeS*	46296/1212	*BS* 168		PksS	P450

Grey KS^0 stands for KS that cannot elongate the polyketide chain. Grey ACP6-1 contains only the N-terminus of ACP.

The location and the module organization of *pks1* (724 kb) is similar to that of the *pksX* from *Bacillus subtilis* 168. The ORFs of *pks1* show similarity between 52% and 83% to *pksX*, except two genes (*pksA* and *pksG*). *pksA*, a hypothetical transcriptional regulator, which locates directly at the beginning of the gene cluster in *Bacillus subtilis* 168, lies 730 kb upstream of the *pks1* gene cluster in FZB42. *pksG*, encoding a β-ketoacyl synthase, is absent in the *pks1* gene cluster. The total 17 modules (Table 5) for PKS and NRPS begin with the gene *baeJ*: three ORFs (*baeL, baeM and baeR*) encode PKS type I modules; two ORFs (*baeJ, baeN*) encode NRPS/PKS hybrid modules. 14 KS domains, 16 ACPs, 9 DHs, 9 KRs, 2 MTs

and 1 TE are involved in the 15 PKS modules of the *pks1* gene cluster (Table 5). No AT was found in any PKS module, except that three AT domains stand freely in front of the gene cluster (BaeCDE). Module 2 and 6 contain two ACPs. The first ACP of module 6 contains only the N-terminal part including the serine for connecting to PPant moiety (Table 5). Except the C-terminus of BaeM and the N-terminus of BaeN, the PKS modules stand separate in two neighbouring proteins. The adenylation domain of the first NRPS module in BaeJ can specifically recognize the amino acid glycine, and the one of the second NRPS module in BaeN can recognize alanine as a specific substrate [130].

Table 6. List of pks2 ORFs and their homologous proteins:

Protein/Gene	Size (Da/bp)	Similar to strain	Similarity [%]	Homolog protein	Domain
MlnA/*mlnA*	86329/2307	FZB42	53	DfnA (pks3)	AT
MlnB/*mlnB*	457737/12261	CC H10	38	beta-ketoacyl synthase	[KS0,KR0,ACP0], [KS1,DH1,KR1, ACP1-1, ACP1-1], [KS2,DH2
MlnC/*mlnC*	179074/4773	CC H10	42	beta-ketoacyl synthase	KR2,ACP2], [KS3,DH3
MlnD/*mlnD*	323718/8709	CC H10	39	beta-ketoacyl synthase	KR3,ACP3], [KS4,KR4,ACP4], [KS5, ACP5-1
MlnE/*mlnE*	258599/7005	BS B11	96	PksB	KR5,ACP5-2], [KS6,KR6,ACP6], [KS7
MlnF/*mlnF*	214038/5712	BS B11	97	PksC	DH7,ACP7-1,ACP7-2,KR7], [KS8
MlnG/*mlnG*	275495/7383	BS B11	97	PksD	KR8,ACP8], [KS9,ACP9-1,KR9, ACP9-2], [KS10
MlnH/*mlnH*	144480/3852	BS B11	96	PksE	DH10,ACP10, KR10, TE]

CC H10 = *Clostridium cellulolyticum* H10; Grey KS^0 stands for KS that cannot elongate the polyketide chain. Grey ACPs are ACPs without serine to connect the PPant moiety.

The *pks2* gene cluster with 53 kb is the smallest *pks* gene cluster in the genome of FZB42. It contains eight ORFs, and no similar gene cluster was found in the genome of *B. subtilis* 168. The *pks2* gene cluster is inserted between *ykyA* and *pdhA*. Except the insertion of the *pks2* gene cluster, the genome sequences of FZB42 and BS 168 are highly conserved in this region. The first gene (*mlnA*) shows 53% similarity to the first gene of *pks3* (*dfnA*), *mlnBCD* have low similarities to the polyketide synthases of *Clostridium cellulolyticum* H10, the other four ORFs of *pks2* are almost identical to the genes present in *Bacillus subtilis* B11 (Table 6). In total, 11 PKS modules are involved in the gene cluster, which include 11 KSs, 15 ACPs, 11 KRs, 5 DHs and 1 TE. Methylase transferase domain could not be detected in the *pks2* gene

cluster (Table 6). Only one AT domain is present in front of the *pks2* gene cluster (MlnA). Modules 1, 5, 7 and 9 contain two ACPs, respectively. One of the ACPs in module 5 and 9 does not provide the serine for connecting to a PPant moiety, so these ACPs cannot convert to the active ACP forms. Except the last module, the PKS modules stand separately in two neighbouring proteins. A unique feature of the *pks2* gene cluster in comparison to the *pks1* and *pks3* is that no cytochrome P450 is involved in the gene cluster of *pks2* (Table 6).

Table 7. List of *pks3* ORFs and their homologous proteins:

Protein/Gene	Size (Da/bp)	Similar to strain	Similarity [%]	Homolog protein	Domain
DfnA/*dfnA*	82951/2259	BS 168	60	PksE	AT
DfnY/*dfnY*	36744/981	n.d.	n.d.	n.d.	n.d.
DfnX/*dfnX*	10026/273	n.d.	n.d.	n.d.	n.d.
DfnB/*dfnB*	49460/1332	*Chloroflexus aurantiacus* J-10-fl	50	AMP-dependent synthetase and ligase	[AL0
DfnC/*dfnC*	26714/738	*Pseudomonas fluorescens*	46	MupS	3-ketoacyl-(ACP) reductase
DfnD/*dfnD*	464004/12591	BAM FZB42 /BS 168	60/46	DfnG/PksP	ACP0],[KS1,DH1,KR1,MT1, ACP1-1,ACP1-2],[KS2, DH2, KR2, ACP2],[KS3
DfnE/*dfnE*	232861/6297	BS 168	48	PksJ	DH3, ACP3],[KS4,KR4, ACP4], [$KS^0$5
DfnF/*dfnF*	212483/5727	*Bacillus subtilis* A1/3	96	PksM	DH5,ACP5],[KS6,ACP6-1,DH6, KR6
DfnG/*dfnG*	573635/15615	BAM FZB42/BS 168	60/38	DfnD/PksN	MT6,ACP6-2,ACP6-3],[KS7,DH7, KR7,ACP7],[KS8,DH8,KR8, ACP8], [KS9,KR9,ACP9], [$KS^0$10
DfnH/*dfnH*	286026/7719	*Geobacter uraniireducens* Rf4	50	beta-ketoacyl synthase	DH10,ACP10,DH10],[KS11, KR11, ACP11],[KS12
DfnI/*dfnI*	225736/6153	BS 168	43	PksN	DH12, ACP12,DH12],[KS13, DH13, KR13
DfnJ/*dfnJ*	230391/6216	*Sorangium cellulosum* 'So ce 56'	47	polyketide synthase	MT13,ACP13,ER13],[KS14, ACP14,ACP14,TE]
DfnK/*dfnK*	44385/1155	*Arthrobacter sp.* FB24	42	cytochrome P450	P450 cytochrome
DfnL/*dfnL*	46216/1248	BAM FZB42	72	BaeG	3-hydroxy-3-methylglutaryl CoA synthase
DfnM/*dfnM*	27475/747	BS168	55	PksI	ECH

Grey KS^0 stands for KS that cannot elongate the polyketide chain. Grey ACPs are ACPs without serine to connect a PPant moiety.
n.d. = Not detected in the NCBI database

The *pks3* gene cluster spans over 195 kb with 15 ORFs and is the lar gest *pks* gene cluster in the genome of FZB42. The *pks3* gene cluster is inserted b etween *yqjM* and *proI* of BS168 by deletion of *yqjN*. Three unique genes, RBA M_022090, RBAM_022100 and RBAM_022110, are inserted upstream of the *pks3* promotor region. RBAM_022090 is a putative transcription anti-terminator, RBAM_022100 is a putative transcription regulator, w hile RBAM_022110 has not been assigned with any functi on. Except for RBAM_022100, the other two unique genes are transcribed in opposite direction to *pks3*. The ORFs in the *pks3* gene cluster show very low similarities to *pks* genes in other species (Table 7). DfnXY did not match any known protein in the NCBI database and did not contain any puta tive conserved dom ain. Fifteen modules are involved in the *pks3* gene cluster containing 15 KS s, 18 ACPs, 11 DHs, 10 KRs, 3MTs, 1 ER and 1 TE (Table 7). L ike *pks2*, *dfnA* contains only one A T domain in the whole gene cluster. Module 1, 6 and 14 contain two ACPs, respectively. One of the ACPs in m odule 6 does not provide the serine fo r connecting to a PPant m oiety. Except the last m odule, the PKS modules stand separated in tw o nei ghbouring proteins. There are three genes *dfnKLM* following the main modules of the pks3 gene cl uster. DfnK contains a conserved dom ain for cytochrome P450, DfnL contains a putative 3- hydroxy-3-methylglutaryl-CoA synthase, and DfnM has a hypothetic enoyl-CoA hydratase (Table 7).
A m ore detailed description of the polyketides gene clusters in FZB42 has been published elsewhere [Lit. 7, Lit. 1 and Lit. 2].

3.2.2.4 Other genes related to nonribosomal synthesis of secondary metabolites

Besides lipopeptides and polyketides, three gene clusters, involved in non-ribosomal synthesis of peptides (bacillibactin, bacilysin and a puta tive peptide), were detected in the genom e of *B. amyloliquefaciens* FZB42.
A gene cluster hom ologous to *Bacillus subtilis* 168 *dhbACEBF* assigned for the synthesis of the siderophore bacillibactin, was detected in the genome of FZB42. Similarities were ranking from 64% to 80% on the am ino a cid level. T he region inclusive this ge ne cluster is well conserved [55]. DhbF encodes a dimodular pep tide synthetase, which contains two modules composed of C, A and PCP domain and a term inal T domain. The substrate specificity of A1 in DhbF is glycine and A2 is threonine.

Table 8. Organisation of the genes in nrs gene cluster:

Protein/Gene	Size (Da/bp)	Similar to strain	Similarity [%]	Homolog protein	Domain
NrsA/*nrsA*	26505/705	*Bacillus cereus* E33L	42	thioesterase	TE
NrsB/*nrsB*	38665/984	*Clostridium beijerinckii* NCIMB 8052	38	conserved hypothetical protein	mcbC-like_oxidoreductase[1]
NrsC/*nrsC*	392167/10236	*Clostridium beijerinckii* NCIMB 8052	45	amino acid adenylation domain	C1,A1,PCP1,C2,A2,PCP2, C3, A3,PCP3
NrsD/*nrsD*	104199/2760	*Clostridium beijerinckii* NCIMB 8052	37	Predicted hybrid NRPS/ PKS	KS, GloB
NrsE/*nrsE*	42983/1098	*Eubacterium dolichum* DSM 3991	27	hypothetical protein	HTH_ARSR, Arsenical Resistance Operon Repressor
NrsF/*nrsF*	62320/1650	*Clostridium beijerinckii* NCIMB 8052	48	Predicted NRPS adenylation domain	Not detected

1. mcbC-like_oxidoreductase is the oxydase domain of NRPS (non-ribosomal peptide synthetase) and other proteins that modify polypeptides by cyclizing a thioester to form a ring.

During annotation, one unique gene cluster was identified, encoding domains for lipopeptide synthetases and polyketide synthases. This gene cluster was named *nrs*. It is located 2,868 kb downstream from the replication origin in counter-clockwise orientation, and is inserted between *ytzC* and *ytqA*. RBAM_027480, located directly in front of the *nrs* gene cluster, is transcribed in its opposite direction. According to the conserved domain analysis, RBAM_027480 is a putative multi-drug resistance gene. The *nrs* gene cluster involves six ORFs and has some unusual features: 1) the thioesterase is not located at the end, but in front of the gene cluster (Table 8); 2) NrsC contains three modules for NRPS. Notably, there are three repeats ranging from the end of C1 up to the middle of T3 spanning a total length of 8982 bp. These repeats show 98% to 100% identity on nucleic acids level. C1 and T3 show a low homology (around 70%) to C2/C3 and T1/T2, respectively (Table 8); 3) three adenylation domains (A1, A2 and A3) possess the same substrate specificity, namely cysteine (Table 8). Until now, the function of this gene cluster could not be elucidated (see later section 3.3.4). *bacABCDE* is responsible for the biosynthesis of bacilysin in different *Bacilli* [101] and is also present in the genome of FZB42, displaying a high similarity (between 84% and 93%) on the protein level to the corresponding gene cluster in *B. subtilis* 168. In comparison with *Bacillus subtilis* 168, the genes in the neighborhood of the *bac* gene cluster are well conserved. A recently published paper claimed that *yvfI* is essential for bacilysin biosynthesis

as a transcription regulator [131]. Corresponding to its localization in BS168, *yvfI* was found about 339 kb upstream of the *bac* gene cluster. LacR, another transcription regulator, which is located upstream of the *bac* gene cluster, is rearranged in the genome of FZB42 and is replaced at its original position by the unique gene RBAM_031480. Analysis of conserved domains revealed that RBAM_031480 contains an aminotransferase. Recently, we found that bacilysin inhibits growth of phytopathogenic *E. amylovara,* the cause of fire blight disease [Lit. 2].

3.2.3 Whole genome screening of drug efflux proteins

Drug efflux proteins can contribute to natural insensitivity to antibiotics and are one of the most important tools used for antibiotics resistance. These proteins are widely distributed in the genome of FZB42 (Supplementary table 1). Total 59 genes, in another words 42 transporter systems, are involved in multidrug transporting, thereof six belong to ABC-transporter systems. Except the ABC-transporter for bacitracin, mersacidin and subtilin, YfiLMN1 is partially similar to BS168, whilst the other two ABC-transporter systems are unique in FZB42 (Supplementary table 1).

Table 9. Features of the proteins belong to SMR (small multidrug resistance) family:

Gene	TMHMM	Protein length [aa]	MW (Da)
ykkC	4	112	11894
ykkD	3	104	11236
ebrA	4	104	11289
ebrB	3	117	12406
yvaE	3	121	13368
yvaD	4	135	16288
yvdS	4	109	12004
yvdR	3	104	11105

SMR (small multidrug resistance) family proteins are the smallest anti-porters that expel various toxicants from bacterial cells to confer multidrug resistance to the cell. They consist of proteins with only 100 to 120 amino acid residues and four transmembrane spanning α-helices [132]. YkkCD and ErbAB, members of the SMR family, could be detected in the FZB42 genome. According to the protein length, two further candidates were found, YvaED and YvdSR. Transmembrane helix analysis showed that YkkC, EbrA, YvaD and YvdS possess four TMHs, but YkkD, EbrB, YvaE and yvdR have only three (Table 9). According to the work from Saiers´ group [133, 134], there is a TMH in the beginning of the N-terminus of EbrB and YkkD, where YkkD, EbrB, YvaE and YvdR share highly conserved sequence

(**Figure 17**). This indicates that YvaE and YvdR have also four TMHs. It was reported that the conserved membrane-embedded glutamate in the first transmembrane helix is absolutely conserved in all SMR family members [135, 136]. Except YvaD, the other seven proteins contain glutamate residues in the first transmembrane helix (Figure 17). The amino acid sequence of the first transmembrane helix from YkkC and YvdS is almost the same, and its glutamate residues are located in the same positions (Figure 17).

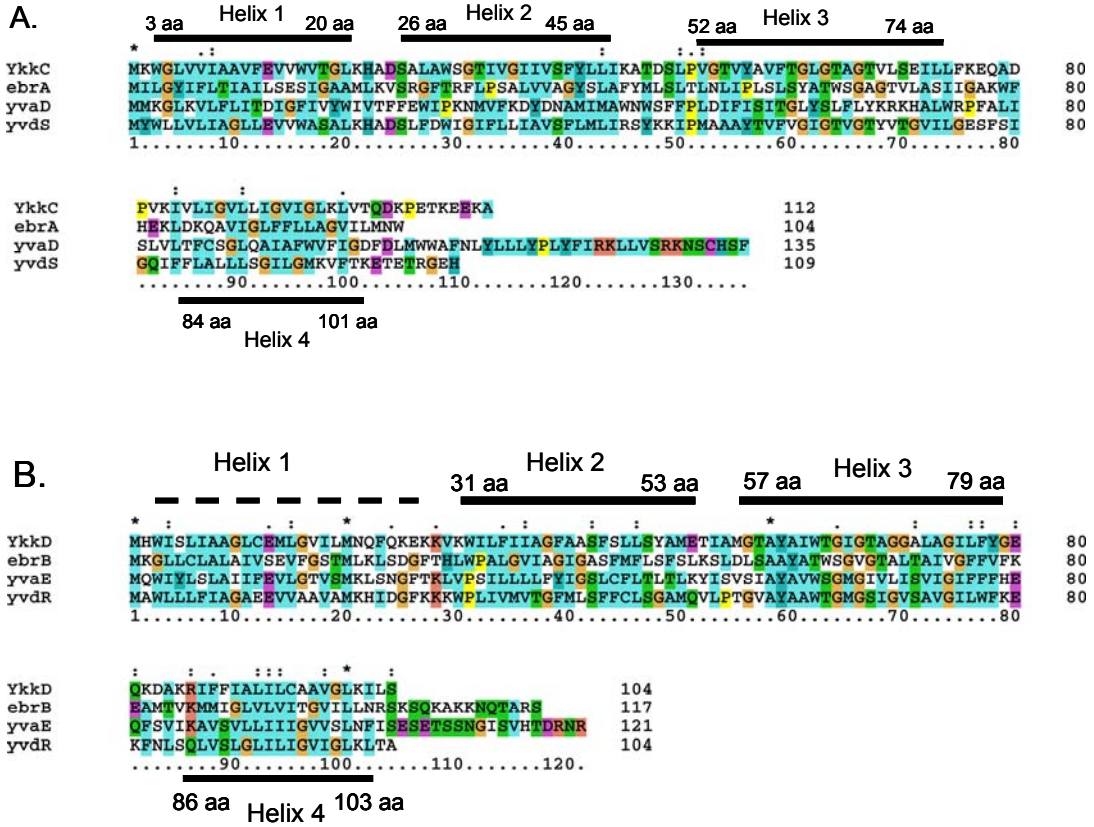


Figure 17 Multiple sequence alignment of putative members of the SMR family: putative transmembrane helix are represented by the lines with the start and stop position on the protein. The solid line stands for the TMH detected by the software TMHMM; the dashed line stands for the TMH prediction by comparison with the work of Saiers` group. A. Proteins with 4 TMHs (YkkC, ErbA, YvaD and YvdS) were aligned. B. Proteins with 3 TMHs (YkkD, ErbB, YvaE and YvdR) were aligned.

3.3 Functional analysis of B. amyloliquefaciens FZB42 genome

3.3.1 Analyzing polyketides produced in *B. amyloliquefaciens* FZB42

In this part, I analyzed genetically the three polyketide gene clusters together with chemical detection methods, like HPLC, LC-ESI MS and MALDI-TOF MS. The results of this analysis have been published in part in several papers (Lit. 1, Lit. 2, Lit. 3, Lit. 4, and Lit. 7). Here I will summarize the most important results and present some additional aspects of my analysis.

3.3.1.1 *Identification of three polyketides produced by* B. amyloliquefaciens *FZB42*

In antibiotic activity tests, our lipopeptide synthetases knoc kout m utants showed decreased antifungal activities at dif ferent levels, but the antibacterial activ ity of FZB42 a gainst *B. megaterium* was not affected (**Figure 18**). This indicates that antibacterial activity of FZB42 is not due to lipopeptides, but m ight be cau sed by polyketides, bac ilysin or other unknown bacteriocins [8]. In order to inv estigate whether the *pks1*, *pks2* and *pks3* gene clusters can produce antibacterial polyketid es, knockout m utants CH6 ($\Delta pks1::cm^R$), CH7 ($\Delta pks2::cm^R$) and CH8 ($\Delta pks3::em^R$) were analyzed by bioautography as well as by chem ical detection methods.

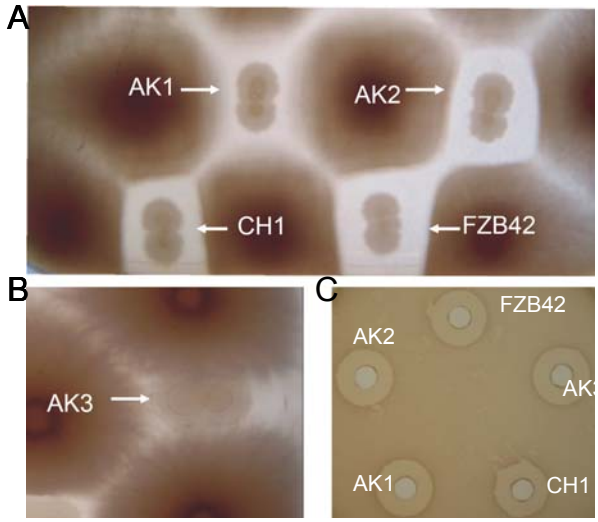


Figure 18. Lipopeptides ar e responsible for antifungal activity: the indicator strain in A and B was Fusarium oxysporum f. sp. cucum erinum DSMZ 62313. I n the single m utants, AK1 (bm y-) shows weak activity ag ainst F. oxysporum , wher eas CH1 (srf-) and AK2 (fen-) have the same effect like wild type FZ B42 on inhi biting gr owth of F. oxysporum. The double m utant (bm y-, fen-) shows loss of almost all of this activity. None of the mutants does suppr ess the gr owth of Bacillus megaterium (C). [Lit. 8]

Bioautography dem onstrated, that in com parison to the wild typ e, in each of the sing le mutants CH6 (*pks1*$^-$), CH7 (*pks2*$^-$) and CH8 (*pks3*$^-$) one of the four inhibition zones produced by FZB42 against *B. megaterium* was m issing (**Figure 19**). The double mutants CH1 1, CH12 and CH14 c orroborated this finding [Lit. 7]. The *pks3* product m igrated slowest among the three polyketides and was located closest to th e start front on the TLC-p late, suggesting that pks3 is m ost hydrophilic. In contrast, pks1 is the m ost hydrophobic com pound due to its fastest m igration in the running buf fer (Figure 19). The high est an tibacterial activity agains t *B. megaterium* was caused by pks3, displaying the bi ggest and clearest inhibition zone on bioautographs (Figure 19). The antibiotic halo produced by pks1 was stronger and more clear than that of pks2, but it seem ed that pks2 did more diffuse into the agar due to its bigger size

of the inhibition zone (**Figure 19**). Among the lipopeptides produced by FZB42, only surfactin showed a weak antibacterial effect (**Figure 19**).

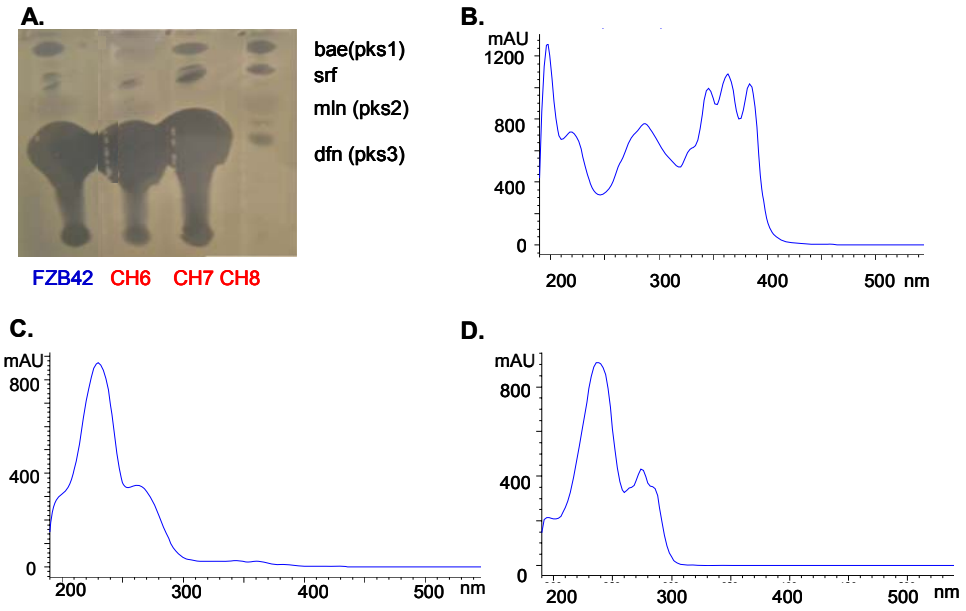


Figure 19 Bioautography of wild type FZB42, CH6, CH7 and CH8 and UV spectrum of the three polyketides: A. Bacillus megaterium was used as the indicator strain; B. UV spectrum of bacillaene; C. UV spectrum of macrolactin; D. UV spectrum of difficidin. The UV spectrum was obtained from190 nm to 550nm.

In cooperation with Professor Süßmuth´s group and Dr. Vater, the three compounds were identified as bacillaene (bae), macrolactin (mln) and difficidin (dfn) using LC-ESI MS and MALDI-TOF MS [113, 119]. This was for the first time that three giant polyketide synthases responsible for biosynthesis of bacillaene, macrolactin and difficidin have been assigned genetically [Lit. 4 and Lit. 7]. The genomic information of these three polyketides allowed to elucidate their biosynthesis pathways (Supplementary figure 1, Supplementary figure 2 and Supplementary figure **3**) and to determine the structure of bae [130]. LC-ESI MS allowed us to identify different derivatives of these three polyketides: six bacillaenes (bacillaene, dihydrobacillaence and another four derivates with unknown modification), four macrolactins (macrolactin A, O- malonyl-macrolactin A, 7- O-succinyl-macrolactin A and macrolactin D) and two molecular forms of difficidins (difficidin and oxydifficidin) were identified in FZB42 (

Table 10). In the following, I will use the terms difficidin, macrolactin and bacillaene for all their derivatives. The retention times and molecular masses of these compounds are summarized in

Table 10.

Table 10 Bacillaene, macrolactin and difficidin and their derivatives produced in B. amyloliquefaciens FZB42:

Polyketide R	R_t (LC-ESI)[1]	R_t (HPLC)[2]	Mass[3]	Modification
Bacillaene (bae 3-A)	n.m.[4]	6.2 (?)	842.41	unknown modification
dihydrobacillaence (bae 3-B)	n.m.[4]	6.3 (?)	844.42	unknown modification
bacillaene (bae 2-A)	6.53	6.5	742.4	hypothetical sugar
dihydrobacillaence (bae 2-B)	6.53	6.5	744.41	hypothetical sugar
Bacillaene (bae 1-A)	7.61	7.40	580.34	basic structure
Dihydrobacillaence (bae 1-B)	7.71	7.51	582.3	basic structure
macrolactin A (mln 1)	7.2	7.26	402.4	basic structure
O-malonyl-macrolactin A (mln 1)	7.39	7.26	488.3	malnoyl-group
7- O-succinyl-macrolactin A (mln 1)	7.43	7.26	502.4	succinyl-group
macrolactin D (mln 2)	6.27	6.011	664.2	succinic acid as a half-ester connected with glucose
difficidin (dfn 1)	10.91	9.041	544.4	basic structure
oxydifficidin (dfn 2)	8.54	8.401	560.3	oxidation

1. Retention time obtained from LC-ESI MS (from the thesis of K. Schneider)
2. Retention time obtained from HPLC
3. Mass obtained from LC-ESI MS (from the thesis of K. Schneider)
4. Not mentioned in the thesis of K. Schneider
5. "?" stands R_t that was not available from LC-ESI MS (from the thesis of K. Schneider)

The HPLC method is commonly used in detecting compounds with UV absorption because of its rapid application, high reproducibility of results and possibility for semi-quantitative determination. Analyzing the wild type and the mutants, at least four bacillaene, two macrolactin and three difficidin isoforms were detected in the supernatant of FZB42 (**Figure 20**). Because HPLC does not allow the determination of the masses of polyketides, their unique UV spectrum was used to identify these polyketides (**Figure 19**B, C and D). However, the different derivatives could not be distinguished in HPLC without their masses. In order to overcome this problem, the retention time was applied in combination with the UV spectrum, because the retention time of these compounds was constant under same conditions. The sample preparation and running condition of HPLC and LC-ESI MS were the same, so masses of the compounds in HPLC were deduced from LC-ESI MS by comparing their retention time (

Table 10). The retention time of different forms of bacillaene was determined as 6.2, 6.5, 7.4 and 7.5 (**Figure 20**). The peak at R_t = 6.2 was attributed to bae 3A and bae 3B. Their modification is unknown. The peak at R_t = 6.5 probably corresponds to bae 2A and bae 2B,

whose modification could be caused by a sugar residue according to their empirical formula. Subject to the area of these peaks, bae 2 is the main bacillaene product of FZB42. The peak at R_t = 7.4 could be assign to bae 1A and R_t = 7.5 to bae 1B. Their structures were determined by Butcher et al. [130]. The retention time of macrolactins was 6 and 7.3 minutes, which indicate that the two isoforms are different in their polarity (**Figure 20**). The peak R_t = 6 was ascertained for mln 2 and R_t = 7.3 for mln 1. The retention times of difficidins were determined with 7.7, 7.9, 8.4 and 9.0 min (**Figure 20**). R_t = 9.01 was supposed to be dfn1 and R_t = 8.4, dfn2. R_t = 7.7 and 7.9 could not be found in LE-ESI MS, so no mass could be assigned to these dfn isoforms.

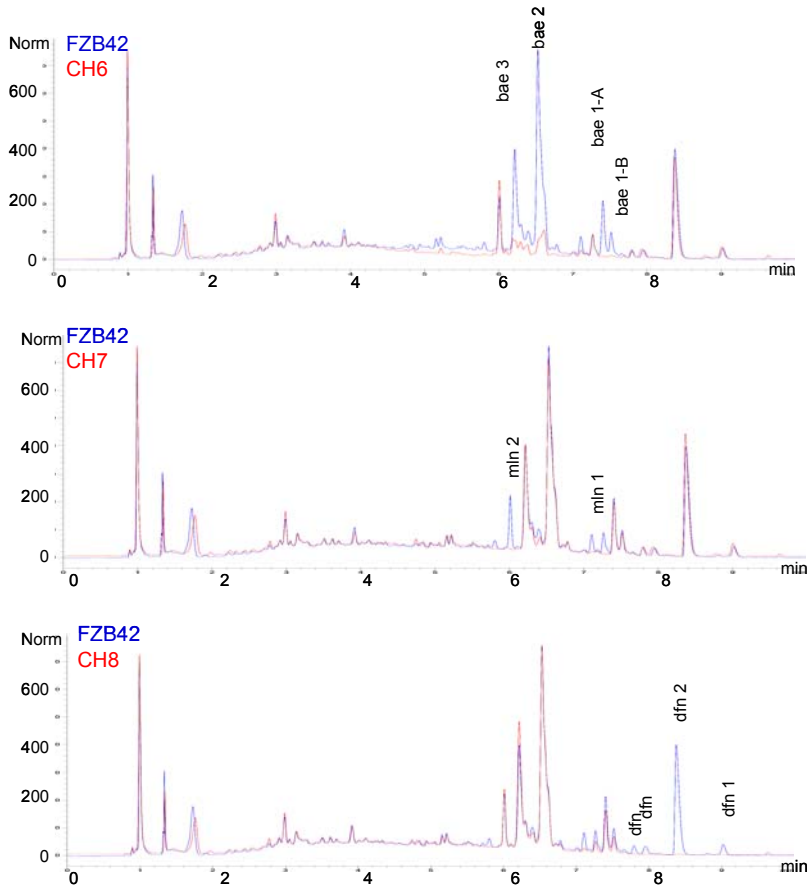


Figure 20 HPLC of the knockout mutants CH6, CH7 and CH8: Four peaks which were disappeared in the HPLC from the supernatants of the mutant CH6 (bae-). In the mln--mutant CH7 two peaks were not present. Four peaks were disappeared in the dfn- -mutant CH8. The chromatographs were selected by the wavelength at 280 nm.

3.3.1.2 Analyzing the roles of trans*-AT in the synthesis of FZB42 polyketides*

One of the most impressed features of the module organization in these three polyketide synthases is absence of ATs in every module. In total, five acyl-transferases (AT) were ascertained within the three gene clusters. Three ATs (Bae1AT in BaeC, Bae2AT in BaeD,

Bae3AT in BaeE) are in front of the *bae* gene cluster, one is upstream of *dfn* (DfnAT in DfnA) and *mln* (MlnAT in MlnA).

To examine the role of these acyl-transferases, the mutants CH24 (Bae1AT$^-$), CH25 (Bae2AT$^-$), CH26 (Bae3AT$^-$), CH27 (MlnAT$^-$) and CH9 (DfnAT$^-$) were constructed by complete deletion of the AT domains. In this work, two approaches were applied in order to analyze the effect of deleting AT domains on biosynthesis of these polyketides.

One of them was to compare the size of the inhibition zones via bioautography. The size of the inhibition activity of a given compound is only dependent on its concentration: the higher the concentration, the bigger the inhibition zone. Deletion of one of the AT domains in the *bae* gene cluster did cause not only disappearance of the respective bae inhibition zone, but also a dramatic reduction of the inhibition zone of dfn. Inhibition of mln was only slightly reduced. On the contrary, knockout of the AT-domain of the *mln* or *dfn* gene cluster did not affect the inhibition zone of bae, only its own inhibition zone was reduced or disappeared (**Figure 21**).

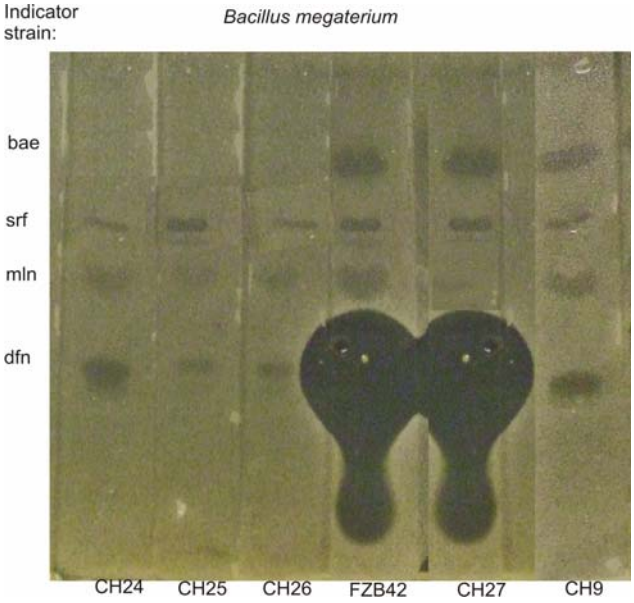
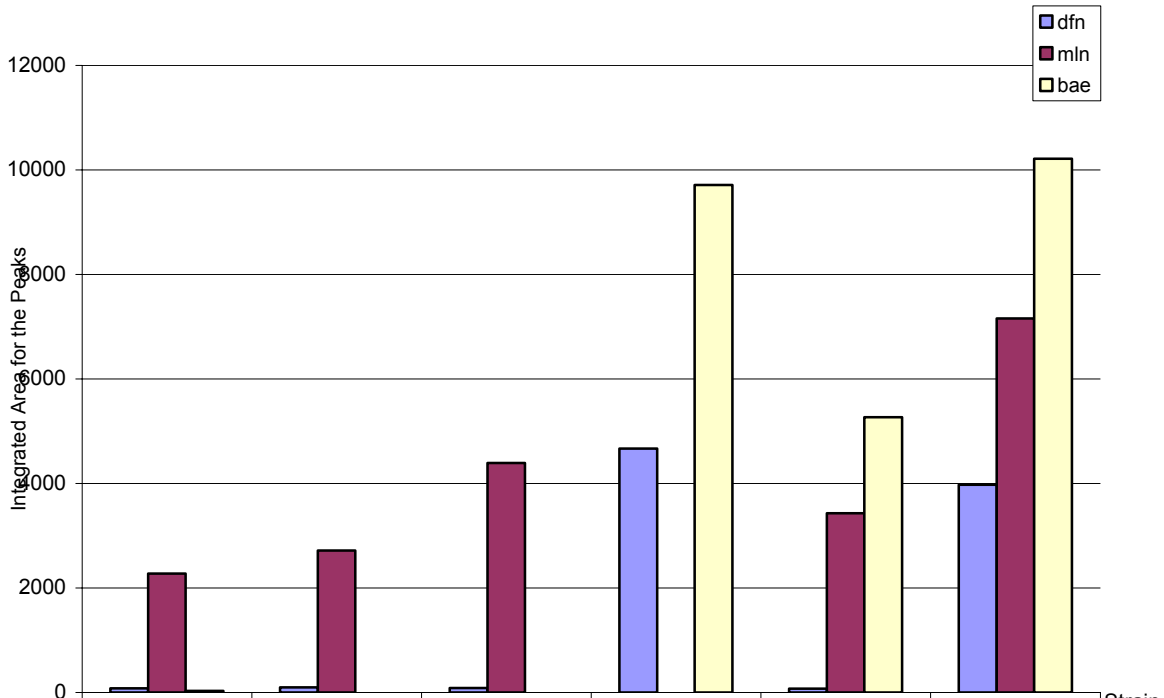


Figure 21 Bioautograph analysis for FZB42 and its AT deletion mutants

Inhibition zones from the bioautography are corresponding roughly to the amount of antibiotic, but this method is quite insensitive. To overcome this disadvantage, a semi-quantitative method coupled with HPLC was also used in analyzing polyketide production. HPLC chromatography of FZB42 and its AT deletion mutants is shown in Supplementary figure **4**. The area of a peak in a chromatograph was determined by integration. Because more than one peak exists for all three polyketides, all peaks from one polyketide were added together. Destruction of BaeAT1, BaeAT2 or BaeAT3 genes with an antibiotic resistance cassette caused: 1) almost complete loss of its ability to produce bacillaenes; 2) strong impact on the

production of difficidins (5-fold reduction in comparison with FZB42 wild type); 3) less effect on the production of macrolactins (1- to 2.5-fold reduction in comparison with FZB42 wild type) (Table 11). Knockout of the AT-domain of the *mln* synthase led to inability to produce macrolactins, whilst the production of difficidins and bacillaenes was not affected (Table 11). Interestingly, deletion of the dfn AT-domain did not only interrupt synthesis of difficidins, but also reduced 2-fold the production of bacillaenes and macrolactins (Table 11). The results from semi-quantitative HPLC analysis corroborated the results from bioautography.

Table 11 Quantification of bae, mln and dfn peaks in FZB42 and their mutants by HPLC:



	CH24 (ΔbaeIAT)	CH25 (ΔbaeIIAT)	CH26 (ΔbaeIIIAT)	CH27 (ΔmlnAT)	CH9 (ΔdfnAT)	wt
dfn	76,7	97,8	86,60	4663,96	73,4	3975,9
mln	2273,5	2716,7	4392,40	0,00	3430,2	7153,4
bae	32	0	0,00	9713,88	5266	10216

3.3.1.3 Analysis of proteins involved in difficidin synthesis

In this work, we have demonstrated for the first time that difficidin is synthesized by the *pks3* (*dfn)* operon. Polyketide synthase of difficidin belongs to type I PKSs. Its biosynthesis pathway follows the co-linearity rule, which allowed to construct a model for the synthesis of difficidin (Supplementary figure **3**, [Lit. 7]). Some genes belong to this operon, but do not encode any module. These are *dfnK*, *dfnL* and *dfnM*. To find out whether these genes are essential for biosynthesis of difficidin, knockout mutants were constructed (**Figure 22**).

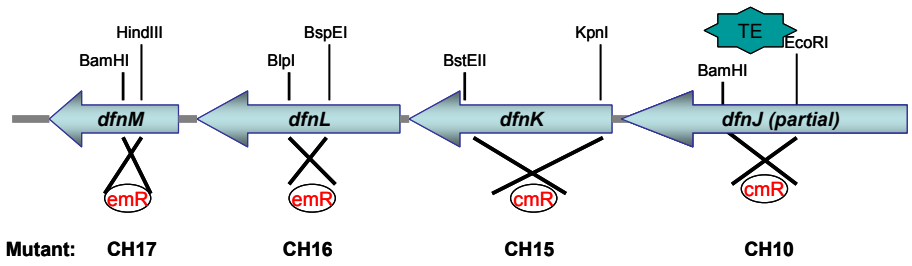


Figure 22 Construction of the mutants in dfn operon.

Protein sequence analysis showed that DfnK possesses a conserved domain (cytochrome P450). Interestingly, BaeS contains also a cytochrome P450 domain, but these two proteins share very low similarity on the amino acids level (about 30%). In the *mln* operon, such a conserved domain does not exist. Comparing the supernatants from FZB42 with CH15, a *dfnK* deletion mutant, revealed that not only difficidin but also macrolactin production decreased almost completely (**Figure 23**). In the case of bacillaene, the results appeared more complex. The production of bae 3 and bae 1 reduced dramatically, however the production of bae 2 increased (**Figure 23**). Interestingly, some new HPLC peaks could be detected in the area between the fourth and sixth minute in the supernatant of CH15 (**Figure 23**), and these peaks did show UV-spectra different from the three polyketides produced by wild-type FZB42.

DfnL contains a conserved 3-hydroxy-3-methylglutaryl CoA synthase domain. Its deletion mutant CH16 was unable to produce difficidins. Production of bacillaenes and macrolactins was not altered. No new peaks could be found in the HPLC from the supernatant of CH16 (**Figure 23**).

The conserved domain in DfnM is an enoyl-CoA hydratase. Deleting *dfnM* (CH17) abolished dfn production completely. The amount of mln 2 (macrolactin D), a dominant macrolactin in the supernatant of FZB42, was reduced in CH 17, whilst the production of macrolactin A and its derivatives remained unchanged. The production of bae 1A, 2 and 3 decreased in the supernatant from CH17, whereas the production of bae 1B remained at the same level as in FZB42 (Figure 23).

It is already known that a thioesterase is essential for the release of the fully synthesized polyketides from PKS [137, 138]. In this work, the function of the TE domain of the *dfn* gene cluster was analyzed by comparison with the TE deletion mutant (CH10). No difficidin was detected in the supernatant of CH10. No intermediates of difficidins could be detected probably due to low sensitivity of our HPLC equipment. The amount of macrolactins remained at the same level as in FZB42, but the bacillaenes production was affected partially,

especially the dominant product bae 2 was reduced dramatically in the supernatant of CH10 (**Figure 23**).

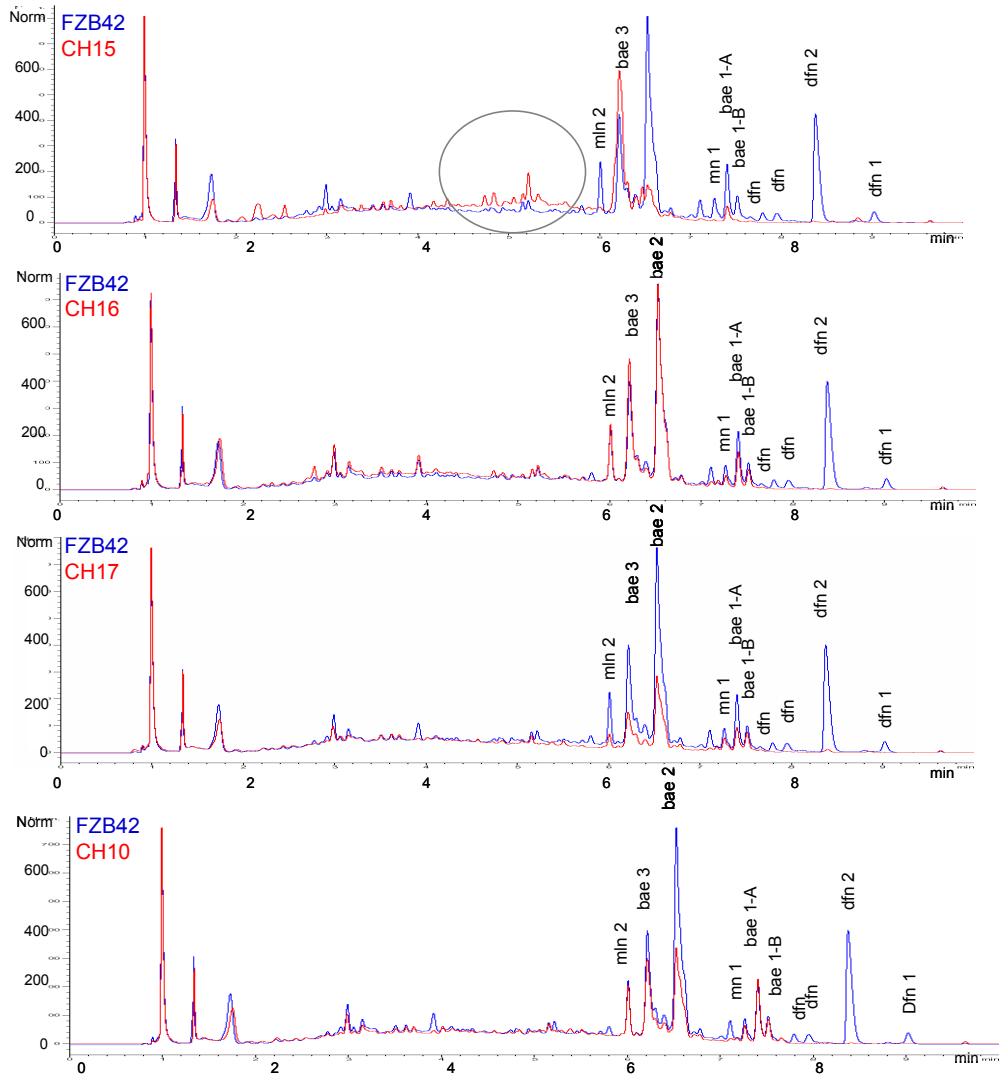


Figure 23. HPLC analysis of CH15 (dfnK-), CH16 (dfnL-), CH17 (dfnM-) and CH10 (TE domain knockout): the polyketides were detected 280 nm; the black circle shows the area of peaks that differ between FZB42 and CH15.

3.3.1.4 Identifying the precursor for the starter and elongation units for macrolactin

To know the substrates that can be utilized in the biosynthesis of macrolactin, a feeding experiment with ^{13}C-labelled sodium acetate was performed. In order to obtain a high quality of ^{13}C-labelling in the macrolactin backbone, the amount of the carbon source, glucose in case of Landy medium, should be kept as low as possible, but cell growth and production of macrolactin should not be strongly affected. To determine the optimal glucose concentration for the growth of CH11, the growth behavior of CH11 in Landy medium with four different glucose concentrations (2%, 1%, 0.5% and 0.25%) was compared. The maximal cell densities were only slightly reduced in decreasing glucose concentrations (Figure 24). Because of their

minor effect on cell growth, Landy medium with the lowest glucose concentration was chosen for the feeding experiment.

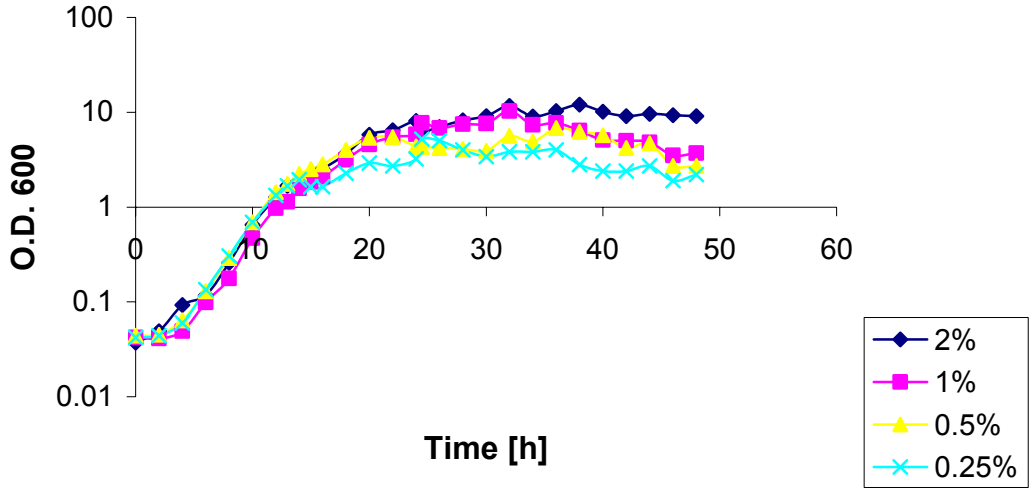


Figure 24. Growth of CH11 in Landy medium with different concentrations of glucose

The labeled macrolactin A was purified from the supernatant by extraction with ethyl acetate and subsequently concentrating by preparative HPLC. The pure compound was analyzed by NMR MS. NMR-spectrum of labeled macrolactin A revealed an alternating labeling of the carbon skeleton with ^{13}C (**Figure 25**). The peaks of twelve ^{13}C-enriched carbon atoms including the starter unit showed similar intensities. No significant isotopic dilution was observed (**Figure 25**), which indicated acetate/malonate as the exclusive precursor for the macrolactin skeleton [Lit. 4].

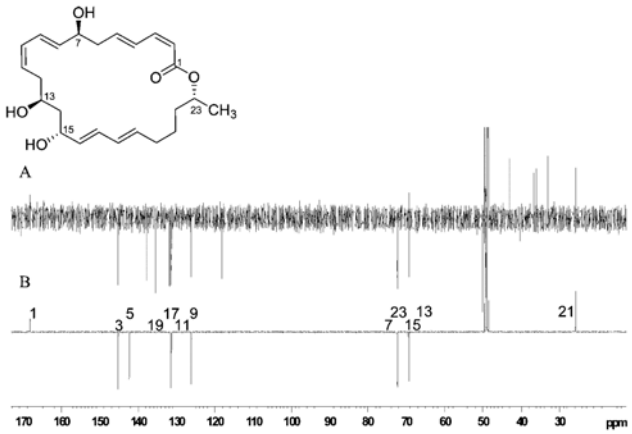


Figure 25. 13C-NMR spectra of (A) macrolactin A and (B) assignment of signals of 13C-labelled macrolactin A derived from [1-13C] acetate feeding experiment: The numbers 1, 3, 5….23 in (B) indicated the C atoms that are labelled by 13C [Lit. 4].

3.3.1.5 Production of bacillaene, macrolactin and difficidin in environmental strains

To understand whether production of the three polyketides is a general feature in *Bacilli,* 29 different *Bacillus* strains were investigated by LC ESI-MS and PCR analysis. The supernatants of these strains grown for 36-hour in Landy medium were extracted by ethyl acetate. Because the AT domains were essential and specific for the production of these compounds (3.3.1.2), five pairs of primers were designed according to the sequences of these AT domains present in the genome of FZB42.

The PCR analysis for the *mln* gene cluster corroborated the LC ESI-MS result, while the results for bacillaene and difficidin were not always corresponding (Table 12).

B. amyloliquefaciens UCMB-5044 and QST713 were found to produce difficidin by LC-ESI MS, but their PCR fragment was bigger than that of FZB42. This indicated that the dfn AT domain is more variable than the one of mln (Table 12).

The HPLC ESI-MS results of the three *B. lichenifo rmis* strains (S167, S173 and B37) indicated that they might produce difficidins, but no PCR fragment of the dfn AT domain could be amplified using chromosomal DNA from those strains (Table 12). According to database and literature, there is no report that *B. licheniformis* can produce difficidins.

The bacillaene gene cluster in the genome of FZB42 contains three different *trans*-ATs. In 14 strains, bacillaene could be detected by both HPLC ESI-MS and PCR with appropriate size, corresponding to the related FZB42 sequence (Table 12). In *B. amyloliquefaciens* DSM7, three bacillaene AT domains were detected by PCR with the same size as in FZB42, but bacillaene could not be detected in the supernatant by HPLC ESI-MS (Table 12). It can be speculated that a dysfunctional PPTase might be the reason for absence of bacillaene in DSM7, as in BS168. In *B. amyloliquefaciens* UCMB-5044 and QST713, bacillaene could be detected by LC-ESI MS, but not all of the three AT domains were amplified by PCR analysis, only the second *baeAT* yielded a PCR fragment corresponding to FZB42. The PCR products obtained from the first and the third *baeAT* were bigger than those of FZB42 (Table 12). In INR937, the first AT domain has the same size as that of FZB42, and the PCR products of the other AT domains were smaller (Table 12). In case of *Bacillus subtilis subtilis* FZB37 and *Bacillus subtilis spizenii* UCMB-5014, bacillaene could be detected by LC-ESI MS in the supernatants (Table 12). The BaeAT specific primers amplified in UCMB-5014 fragments with different sizes to that of FZB42. No PCR products could be detected in FZB37 when using the primers for BaeATs (Table 12). Taken together, these results suggest that the sequences of AT domains in the *bae* gene clusters are highly variable.

Table 12. Production of bacillaene, macrolactin and difficidin in different Bacillus strains detected by the PCR products of the AT domains and HPLC ESI-MS:

strain	Macrolactin PCR	Macrolactin MS	Bacillaene PCR 1. baeAT	Bacillaene PCR 2. baeAT	Bacillaene PCR 3. baeAT	Bacillaene MS	Difficidin PCR	Difficidin MS
	mlnAT	MS	baeAT	baeAT	baeAT	MS	dfnAT	MS
Bacillus amyloliquefaciens B9601-Y2	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 24	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 42	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 45	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 109 (A1/3)	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 113 (QST 113, Serenade)	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FZB 111 (GB03, Kodiak)	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens GaoB2	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens GaoB3	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens Gao55	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens Gao85	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens GB 03 (Kodiak)	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens UCMB-5044	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens UCMB-5113	+	+	+	+	+	+	+	+
Bacillus amyloliquefaciens FDK21	+	+	>	+	>	+	>	+
Bacillus amyloliquefaciens QST713 (Serenade)	+	+	>	+	>	+	>	+
Bacillus amyloliquefaciens DSM7	-	-	+	+	+	-	-	-
Bacillus amyloliquefaciens BE 20-2	-	-	-	-	-	-	-	-
Bacilus licheniformis ATCC 9789 (B27)	-	-	-	-	>	-	-	-
Bacilus licheniformis (B37)	-	-	-	-	>	-	-	+
Bacilus licheniformis S167 (9945A)	-	-	-	-	-	-	-	+
Bacilus licheniformis S173 (Gibson 1158)	-	-	-	-	>	-	-	+
Bacilus licheniformis S175 (NCIMB 7724)	-	-	-	-	>	-	-	-
Bacillus mojavensis UCMB-5075	-	-	-	<	+	-	-	-
Bacillus mojavensis MB2	-	-	-	<	-	+	-	-
Bacillus pumilus INR-7	-	-	-	-	-	-	-	-
Bacillus subtilis spizenii At3	-	-	<	<,>	<,>	-	-	-
Bacillus subtilis spizenii UCMB-5014	-	-	<	<,>	<	+	-	-
Bacillus subtilis subtilis INR 937	-	-	+	<	<	+	-	-
Bacillus subtilis subtilis FZB 37	-	-	-	-	-	+	-	-

"+" stands for presentation of the compound in HLC-ESI MS or the PCR fragment with the same size like FZB42;
"-"stands for absence of the compound in LC-ESI MS or the PCR fragment;
">" stands for the PCR fragment was bigger than the one of FZB42; "<" stands for the PCR fragment smaller than the one of FZB42;
Blue = the result from PCR agreed with the result from LC-ESI MS and the size of the PCR fragment was same to the one from FZB42;
Yellow = PCR fragment with the size like FZB42, but no corresponding compound in LC-ESI MS;
Pink = LC-ESI MS could detected the compounds, but the PCR fragment was absent or present in a size differing to that of FZB42.

In summ ary, all plant growth prom oting *B. amyloliqu efaciens* (16 strains) can produce bacillaene, m acrolactin and dif ficidin [139]. The two *B. amyloliquefaciens* without PGPR effect can produce none of these polyketides. The four strains that can only produce bacillaene belong to *Bacillus subtilis* (three strains) or *Bacillus maja vensis* (one strain), and their AT domains varied to the ones from FZB42 strongly (Table 12).

3.3.2 Role of *yczE* in production of secondary metabolites

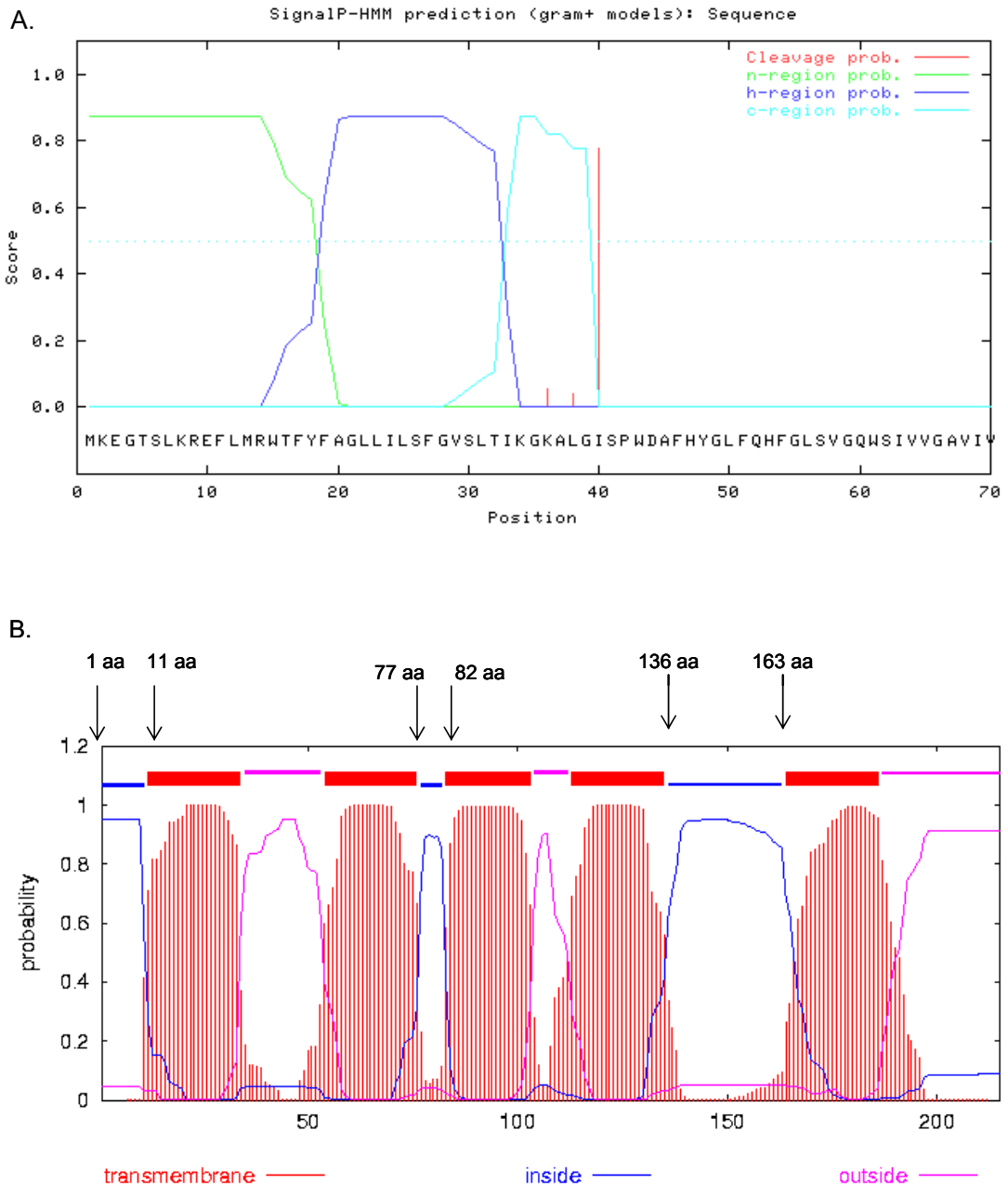


Figure 26. Com putational analysis of the pr otein YczE: A. YczE contains a possible signal peptide; B. YczE is a hypothetical trans-membrane spanning protein.

yczE (215 aa) is located directly downstream of *sfp* and is transcribed in opposite direction to DNA replication and *sfp* transcription. YczE has a signal pe ptide that could be cleaved of f

between 39 aa and 40 aa of the protein with five trans-membrane helices (Figure 26). The main part of the YczE protein consists of trans-membrane helices whilst cytoplasmic and extracytoplasmic segments are very short. No functional domain could be found in this protein.

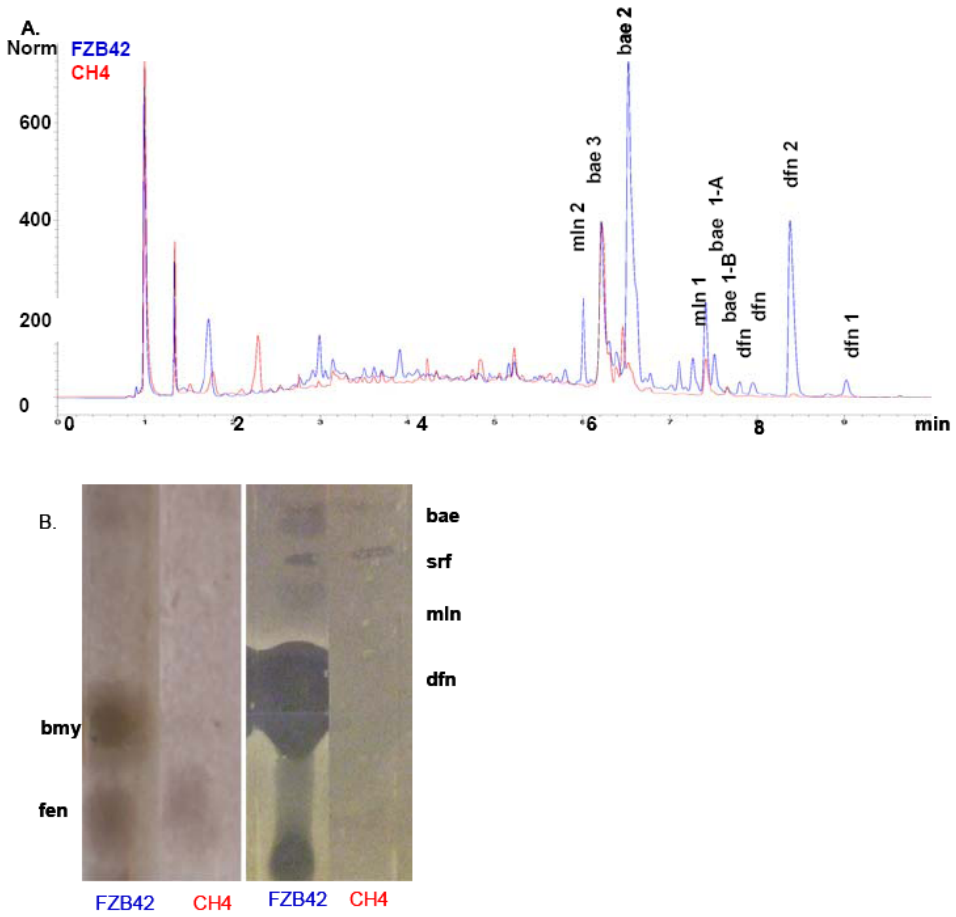


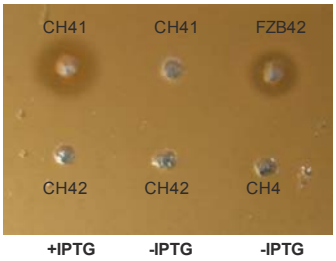
Figure 27. Effect of YczE on production of polyketides and lipopeptides: A. HPLC shows reduction in production of bacillaene (bae), macrolactin (mln) and difficidin. B. Loss of antibiotic activity of the polyketides and surfactin (srf) against Bacillus megaterium (right); loss of antibiotic activity of bacillomycin D and (partially) fengycin (fen) against Fusarium oxysporum (left).

During genome annotation, it was noticed that YczE hits to YczE3B (99% similarity at the amino acid level). *yczE3B* is in the region where the sequence is quite conserved in many surfactin producing *Bacillus* strains, so it was thought that *yczE3B* could be involved in production of surfactin, but no experimental evidence was provided [140]. To verify this hypothesis, *yczE* was disrupted in the genome of FZB42 by insertion with an erythromycin resistance cassette. The production of lipopeptides and polyketides was examined by bio-autography, HPLC and MALDI-TOF MS. Surprisingly, bioautography demonstrated that the inhibition zone of surfactin against *B. megaterium* remained in CH4 (*yczE*$^-$), but the activity of bacillaene, difficidin and macrolactin disappeared. Bacillomycin D, the most prominent

lipopeptide produced by FZB42 in Landy medium, lost its activity against *Fusarium oxysporum* in CH4, whilst production of fengycin was only slightly reduced in CH4 (Figure **27**B). Using HPLC, the amount of the polyketides was estimated. The production of macrolactin and difficidin was completely abolished in CH4. However, production of bae in CH4 was only partially affected: production of bae 3 remained constant whilst production of bae 2 and bae 1 was reduced significantly (Figure **27** A).

To confirm whether YczE is involved in regulation of lipopeptides and polyketides production on transcriptional or translational level, the promoter fusion strains of FZB42 (CH41) and CH4 (CH42) whose original promoter of the *bmy* cluster was replaced by the promoter P_{spac} from pECE149. The *lacI* gene was co-transformed in CH41 and CH42 by insertion in front of the P_{spac} promoter, so that the expression of bacillomycin D could be induced by IPTG and became independent of its possible transcriptional regulators. *Saccharomyces cerevisiae* was sensitive only against bacillomycin D among all antibiotics produced by FZB42, and was used as an indicator for the agar-diffusion test indicating the production of Bacillomycin D in CH41 and CH42. 30-µl supernatants from CH41 and CH42, both with or without IPTG, were added to check the activity of Bacillomycin D. CH4 was used as a negative control and FZB42 as a positive control. The bioassay showed that CH41 could restore its ability to produce bacillomycin D after adding IPTG, and implied that the promoter P_{spac} in front of the *bmy* gene cluster was functional. In contrast, CH42 did not restore the ability to produce bacillomycin D even after adding IPTG, which indicated that YczE could not be a transcriptional regulator of the *bmy* gene cluster.

A.



B.

	FZB42	CH42/ IPTG-	CH42/ IPTG+	CH4	CH4/ IPTG-	CH4/ IPTG+
Bmy synthetases	+	-	+	+	-	+
YczE	+	+	+	-	-	-
Antibiotic activity	+	-	+	-	-	-

Figure 28. Bacillomycin D antibiotic activity in the strains whose bmy gene cluster was regulated by IPTG: A. Saccharomyces cerevisiae was used as an indicator. FZB42 and CH4 were used as controls grown in Landy medium without IPTG [Lit. 5]. B. YczE affects bmyD synthesis

To confirm whether YczE is involved in regulation of lipopeptide and polyketide synthesis, promoter fusion strains of FZB42 (CH41) and CH4 (CH42) whose original promotor of *bmy* cluster was substituted by the Pspac promoter from pECE149. The *lacI* gene was integrated into CH41 and CH42 in front of the Pspac promoter, resulting in IPTG-inducible expression

of bacillomycin D independently from possible transcriptional regulators. *Saccharomyces cerevisiae* was sensitive only against bacillomycin D among all antibiotics produced by FZB42, and was chosen as indicator for production of Bacillomycin D in CH41 and CH42. 30-µl supernatants from CH41 and CH42, cultivated with or without IPTG, were used to determine activity of Bacillomycin D. CH4 was used as negative control and FZB42 as positive control. The agar diffusion-test showed that CH41 can restore its ability to produce bacillomycin D by adding IPTG, and implied that Pspac promotor in front of the *bmy* gene cluster was functional (Figure 28). In contrast, CH42 could not restore ability to produce bacillomycin D even by adding IPTG, which indicated that YczE is not a transcriptional regulator of the *bmy* gene cluster promoter (Figure 28, Lit. 5).

3.3.3 Role of *sfp* in the production of secondary metabolites

Two 4′-phosphopantetheinyl (Ppant) transferase genes are present in the genome of FZB42, *sfp* and *acpS*. AcpS is known to be involved in production of fatty acids, and Sfp is essential for synthesis of active NRPS. Similar to BS168, *sfp* is located downstream from the *srf* gene cluster. However, instead of *ycxA* and *ycxB*, two unique genes, RBAM_003700 and RBAM_003710 are present between *sfp* and *srf* in the genome of FZB42. RBAM_003700 contains a conserved domain with similarity to aminotransferase, whereas RBAM_003710 possess no conserved domains. No other features useful to predict a possible function of these two unique genes do exist.

It is known that different 4′-phosphopantetheinyl transferases share only low similarity (10% to 20%) [95]. Therefore, it is very difficult to identify alternative Ppant transferase in the genome of FZB42 by BLAST. Sfp can convert *apo*-ACPs to *holo*-ACPs *in vitro*, but no investigation has been carried out to evaluate the function of Sfp in context of a real cellular system. In order to investigate roles of Sfp in FZB42, *sfp* gene was disrupted in FZB42. The supernatant of CH3 (*sfp*$^-$) grown in Landy medium was analyzed by bioautography and HPLC. CH3 did not develop inhibitory activity against *B. megaterium* and *F. oxysporum* (Figure 29B). HPLC confirmed that neither polyketides nor lipopeptides were produced by the CH3 mutant (Figure 29A). By contrast, the *sfp*^{0} *Bacillus subtilis* can produce both surfactin and bacillaene after restoring the functional Sfp [Lit. 7].

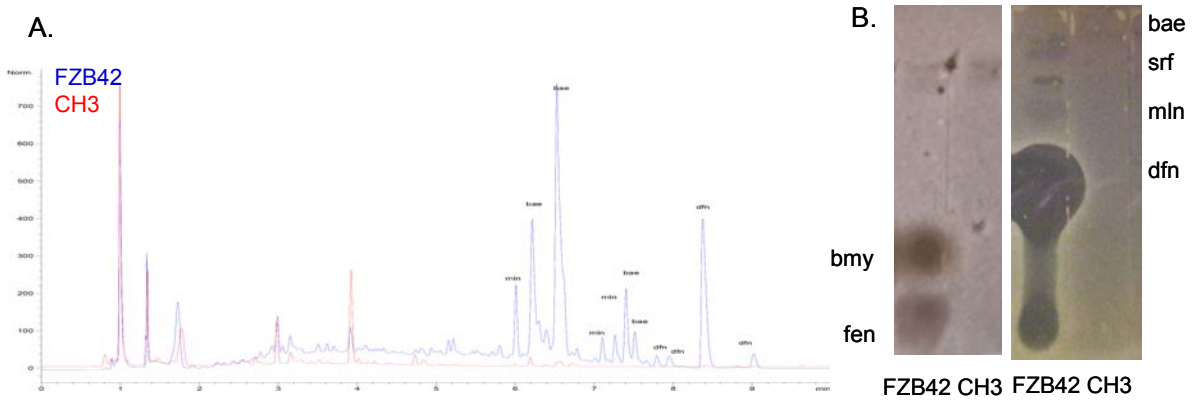


Figure 29 Effect of Sfp on production of polyketides and lipopeptides: A. HPLC showed no production of bacillaene (bae), macrolactin (mln) and difficidin. B. Bioautography: loss of antibiotic activity of polyketides and surfactin (srf) against Bacillus megaterium (right); loss of the antibiotic activity of bacillomycin D and Fengycin (fen) againstFusarium oxysporum (left)

3.3.4 The *nrs* gene cluster

It was not known so far whether the *nrs* gene cluster is expressed in FZB42. To find out the conditions for the transcription of the *nrs* gene cluster, reverse transcription PCR (RT-PCR) was performed. FZB42 cells grown in Landy medium were collected at mid-exponential phase for isolation of total RNA. The purity of RNA was checked by PCR using the primer pair nrs2f and nrs1-r. In order to analyze expression of the *nrs* gene cluster, the primer nrs2f (5´-TGGTATGCTAACATATAATCTC-3´) was used for reverse RT-PCR. As a negative control, the primer nrs2-r was used. To visualize the cDNA product on agarose gel, the cDNA was amplified by the primer pair nrs2-f and nrs1-r. A 273-bp fragment was obtained, if the primer nrs2-f was used, but no PCR product could be obtained in the case the primer nrs2-r was used. This result demonstrated that the *nrs* gene cluster is transcribed in mid-exponential phase (Figure 30).

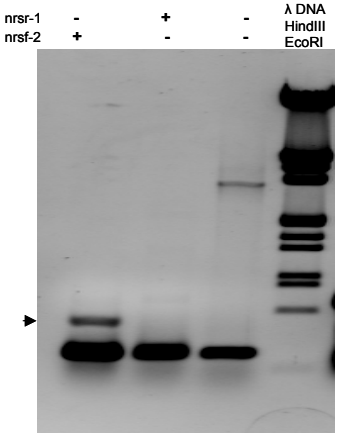
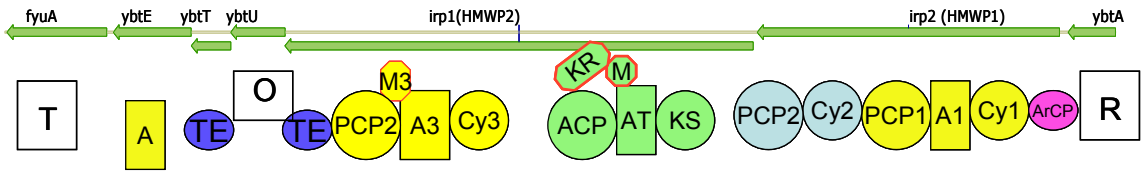


Figure 30. RT-PCR product from the nrs gene cluster

The *nrs* gene cluster involved three modules for the lipopeptide synthesis, which is responsible for condensation of three cysteines. The module organization is very similar to that of yersiniabactin synthetase [141, 142]. The gene cluster encoding yersiniabactin synthetase contains five genes: *irp2*, the first gene of the operon, contains two modules of NRPS for two cysteines; *ipr1* contains one PKS module and one module for NRPS for the third cysteine, while *nrsD* contains three NRPS modules for cysteine; the condensation domains of these cysteine can circularize the cysteine itself (**Figure 31**). A transcriptional regulator and a transporter gene were found adjacent to the structural genes (**Figure 31**). Similar to yersiniabactin synthetase, the *nrs* operon contains also three cysteine modules and their condensation domains can circularize cysteines. NrsA and NrsB contain two thioesterases. However, it is unusual in NRPS or PKS systems that thioesterases are located in front of the gene cluster (**Figure 31**). Unlike yersiniabactin, these three modules are located in one protein, NrsC (**Figure 31**). Instead in the centre of the NRPS modules, the PKS module is located downstream of the NRPS modules. ArCP, the starter unit of yersiniabactin synthetase, is not present in the *nrs* gene cluster. NrsE contains the conserved domain CD00090, a metal regulated repressor. RBAM_02748 is located upstream of the *nrs* promoter, its N-terminus contains a conserved domain for multi-drug efflux (**Figure 31**). However, at its C-terminus no conserved domains could be detected. Similarity in organization of the modules suggests that nrs might have siderophore function.

A. Gene cluster for yersiniabactin biosynthesis



B. *nrs* gene cluster for biosynthesis of unknown compound

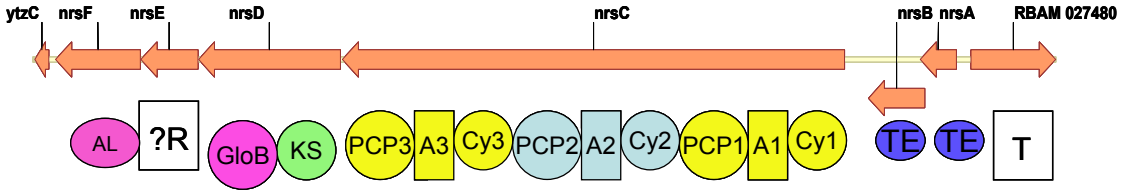


Figure 31. Module organization of the nrs gene cluster in comparison with the yersiniabactin biosynthesis gene cluster: A=adenylation domain A, AL= acyl Co-A synthetase, ArCP= aryl carrier protein, Cy=cyclization, PCP=peptidyl carrier protein, KS= β-ketoacyl synthase, AT=acyltransferase, M= methyltransferase, KR=ketoreductase, ACP=acyl carrier protein, O=Oxidoreductase, R= transcriptionalregulator, T= transporter.

As earlier mentioned, a *dhb* gene cluster exists in the genome of FZB42 and is responsible for production of the siderophore bacillibactin (3.2.2.4). The mutant strains, CH38 with deletion of the whole *nrs* operon, CH39 with deletion of the whole *dhb* operon and CH40 with complete deletion of both, *dhb* and *nrs* gene clusters, were constructed in order to evaluate function of nrs in iron transport. FZB42 and its mutants (CH38, CH39 and CH40) were grown in Landy medium with different iron concentrations (0.1 µM, 0.2 µM and 0.4 µM Fe_2O_3) at 37°C and 210 rpm. The cell density was measured at $O.D._{600}$ every two hours for 13 hours. CH38 (*nrs*$^-$) had a similar cell density as the wild type. CH39 (*dhb*$^-$) and CH40 (*nrs*$^-$ and *dhb*$^-$) grew more slowly than FZB42, but no obvious differences in cell growths were observed between CH39 and CH40 (**Figure 32**).

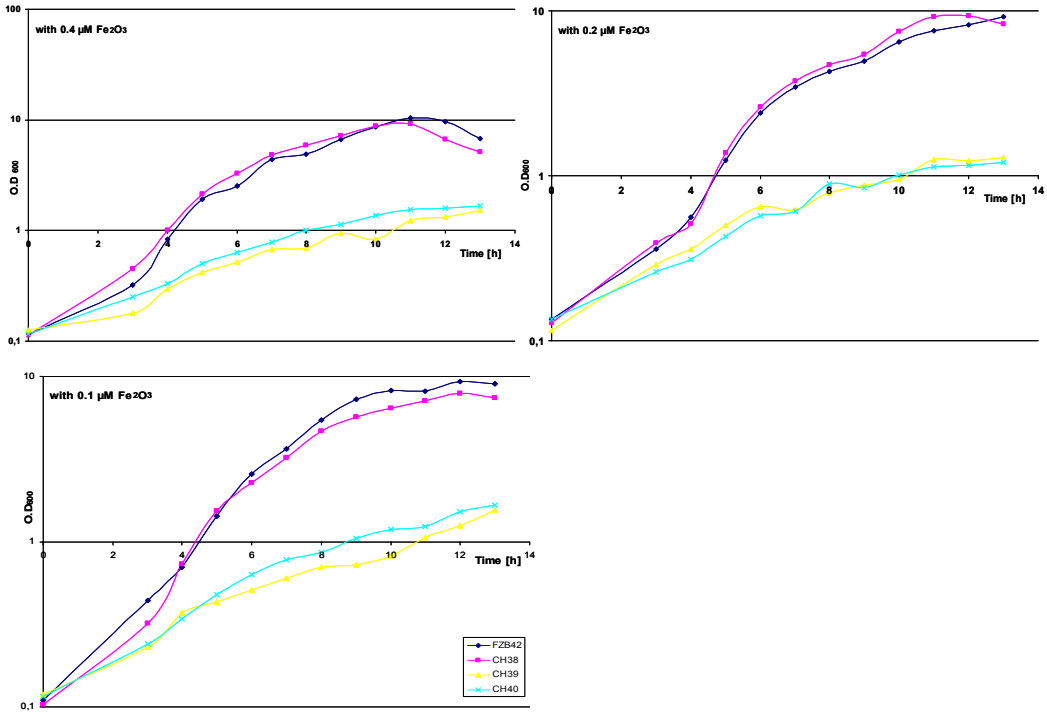
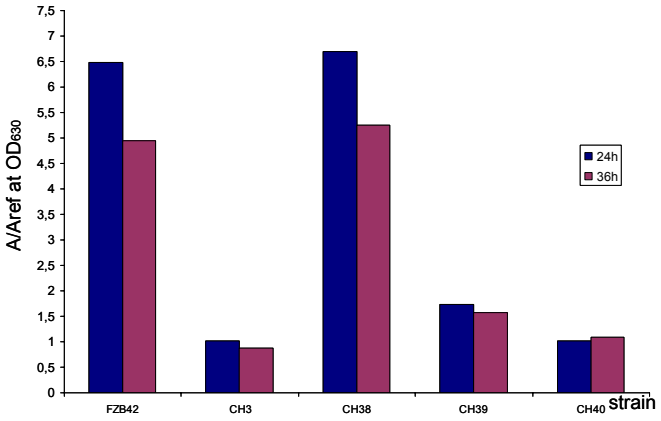


Figure 32. Cell growth curves of FZB42, CH38, CH39 and CH40 in Landy medium with different iron concentrations.

The iron deprivation experiment did not reveal whether nrs belongs to the siderophore family. Therefore, I performed an additional experiment. CAS-assay was performed with supernatants from cell cultures grown in Landy medium. In order to eliminate the background from unspecific iron complexes, siderophore-free CH3 was used as negative control. FZB42 and CH38 secreted the highest amount of siderophores after 24-hours of growth. Concentration of siderophores decreased after 36-hours. Whilst CH38 (*nrs*$^-$) and wild type FZB42 did not differ, the amount of siderophores was dramatically reduced in CH39 (*dhb*$^-$)

after 24 and 36 hrs growth. These results indica ted that siderophores are produced in FZB42 and the *nrs* mutant strain, but nearly nothing in CH39. Notably, the amount of siderophores in CH39 was slightly hi gher than in the *sfp* mutant strain CH3 and the double *dhb nrs* mutant CH40 (Figure 33) suggesting that nrs might somehow contribute to siderophore production in FZB42.

In order to identify the product(s) of the *nrs* gene cluster, HPLC, LC- ESI MS and MALDI - TOF MS were used to detect com pounds in the supernatants of FZB42, CH38, CH39 and CH40 grown in Landy medium for three days, however, no noticeable peak could be detected in any of these strains. In addition, cell surface extraction coupled with MALDI-T OF MS, which allows detection of compounds binding at the cell membrane, did not yield results.



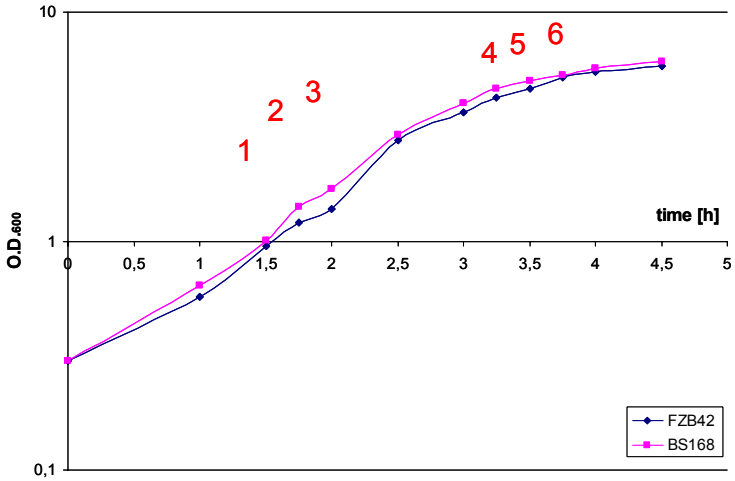
	24h	36h
FZB42	6,4865	4,955
CH3	1,0135	0,8784
CH38	6,6892	5,248
CH39	1,7342	1,57658
CH40	1,0135	1,0919

Figure 33. CAS assay: Using the supernatants of FZ B42 and its mutants in Landy medium with 0.1 µM Fe2O3

3.3.5 Competence and transformation in Bacillus *amyloliquefaciens* FZB42

To find out whether *B. amyloliq uefaciens* FZB42 is transform able under laboratory conditions, FZB42 was grown in GCHE medium and cells were harvested at dif ferent times from m id-exponential growth phase to late st ationary phase. The experim ents showed that transformation happened in the m id-exponential phase and the time-span for competence was very short. However, it was not clear whet her transform ation during m id-exponential phase resulted from a feature unique for FZB42 or was dependent on the transform ation medium (GCHE) that differs from that used in Spizizen´s method [104].

A.



B.

time [h]	FZB42 (OD600)	BS168 (OD600)
1.5	0.95	1.01
1.75	1.2	1.42
2	1.38	1.7
3.25	4.22	4.65
3.5	4.65	5.03
3.75	5.17	5.88

Figure 34. Cell growth in GCHE medium: A. cell growth curves; B. the red labelled points (1, 2, 3, 4, 5 and 6) were gathered for investigating transformation rate.

In order to answer this question, development of competence in FZB42 was compared with that of BS 168. Linearized plasmids pFB01 and pDG268 (1 µg/ml) were used as exogenous DNA for FZB42 and BS168, respectively. The reason for using these two integration vectors was that both possess similar features. An erythromycin resistance cassette was inserted into the *amyC* gene from FZB42 using plasmid pFB01. Correspondingly, a chloramphenicol resistance cassette was inserted into the homologous *amy* gene of BS168 using plasmid pDG268. The lengths of the homologous flanking *amy* sequences from FZB42 and BS168 in pFB01 and pDG168 were around 900 bp each. The overnight LB cultures of FZB42 and BS168 were diluted into MDCH at O.D. $_{600}$ of 0.3. The cells were harvested at three different times during mid-exponential and transient phase. True transformants could grow on agar with the corresponding antibiotics (erythromycin or chloramphenicol). They were unable to hydrolyze starch as indicated by absence of clearing zones onto starch containing agar due to disruption of the *amyC* gene. Efficiency of transformation was calculated by the ratio of the number of transformants and the total number of living cells. The growth curve of FZB42 and BS168 showed two plateaus, one was in the mid-exponential phase and the other in the transient phase (**Figure 34** A). Notably, FZB42 was only competent during the plateau of the mid-exponential phase with a low transformation rate (between 6E-7 to 35E-7) (Figure 35). On the contrary, only one transformant was found in the mid-exponential phase of BS168 and by then, transformation rate of BS tremendously increased during transient phase (Figure 35). The time-window in which FZB42 was competent was found surprisingly narrow, because the

transformation rate reached its maximum at approximately O.D. $_{600}$ 1.2 and decreased significantly either at 0.98 or at 1.38. In contrast, the time window for competent BS168 was wider and the transformation rate did little change during the 30-min period of investigation.

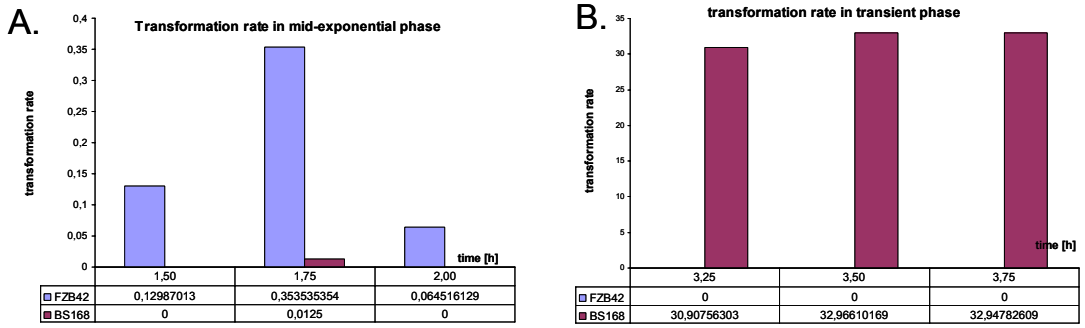


Figure 35. Transformation rate of FZB42 and BS168 in GCHE medium: A. at the mid-exponential phase; B. at the transient phase.

4 Discussion

4.1 Genomic analysis of B. amyloliquefaciens FZB42 give deep insights into the mechanisms for plant growth promotion in Gram-positive bacteria

As a representative of plant associated *B. amyloliquefaciens*, the genome of environmental FZB42 possess versatile gene clusters involved in plant growth promotion and biocontrol. Details about these genes are described in our previous publications (Lit. 6, Lit. 7 and Lit. 8) and reviews (Lit. 1 and Lit. 2). In this chapter, several molecular mechanisms exerted by FZB42 to enhance the plant growth are summarized.

(1) ***Biofertilization:*** it is common knowledge, that not the absolute amount of the nutrients, but the availability of these nutrients in the soil is the limiting factor for plant growth. FZB42 can improve the nutrient availability in soil by different ways. Phytase, encoded in *phyC*, can be secreted by FZB42 into the environment, when concentration of phosphate is low [16, 143]. Phytase converts insoluble organic phytic acid into the inorganic form available for plants. Furthermore, FZB42 secretes numerous metal binding proteins that can form soluble chelate complexes with metal ions, like copper, iron, molybdenum. Secretion of siderophores, like bacillibactin is just one example. The *dhb* gene cluster directs nonribosomal synthesis of bacillibactin. It was reported that bacillibactin binds soluble Fe^{3+} to form a soluble siderophore-iron complex that can be actively transported into plant cells [144].

(2) ***Phytostimulation:*** Phytostimulators enhance plant growth in a direct way. The mechanisms behind the stimulatory effect on root development and crop yield caused by FZB42 we are beginning to understand on the base of the genomic information available. Besides improvement of nutrients availability, FZB42 can also produce the phytohormone indole-3-acetic acid (IAA) to stimulate plant growth [18]. A careful examination of the whole genome sequence of *B. amyloliquefaciens* FZB42 revealed three candidate genes with apparent homology to genes previously reported to be involved in IAA metabolism: *ysnE,* encoding a protein similar to IAA acetyltransferase from *Azospirillum brasilense* [145], *dhaS,* similar to indole-3-acetaldehyde dehydrogenase from *Ustilago maydalis* [146], and *yhcX* encoding a putative nitrilase similar to nitrilase2 from *Arabidopsis thaliana* [147]. Disruption of these genes performed in our group confirmed that YsnE and YhcX, but not DhaS, might participate in tryptophan-dependent IAA biosynthesis in FZB42 [18].

(3) ***Biocontrol:*** plant diseases are usually caused by phytopathogens, which can derogate plant growth. FZB42 protects plants from pathogenic bacteria and fungi by producing various antibiotics. The gene clusters, *srf, bmy, fen*, in total covering 137 kb, direct non-ribosomal

synthesis of the mainly antifungal acting lipopeptides surfactin, bacillomycin and fengycin (3.2.2.2). The gene clusters *mln, bae*, and *dfn*, with a total length of 199 kb were shown to direct synthesis of the antibacterial acting polyketides macrolactin, bacillaene, and difficidin. In addition, another operon encodes proteins involved in synthesis and export of the antibacterial acting dipeptide bacilysin. Recently, we have found that not only difficidin but also bacilysin play a key role in attacking fire blight, a serious disease of orchard trees caused by the plant pathogenic bacterium *Erwinia amylovora* [Lit. 2].

On the contrary, gene clusters involved in ribosome dependent synthesis of lantibiotics are less represented in the FZB42 genome. Apart of two incomplete gene clusters involved in synthesis of mersacidine and subtilin, only one peptide-like compound especially efficient against *B. subtilis* mutant strains deficient in the alternative sigma factor ECF W has been detected in the culture fluid. The impressive genetic potential to produce antagonistic acting secondary metabolites, not only enables FZB42 to cope successfully with competing organisms within its natural environment, but also protects the plants from pathogenic bacteria and fungi.

(4) ***Induced systemic resistance (ISR) mediated by rhizobacteria:*** Plant growth-promoting rhizobacteria, in association with plant roots, can trigger induced systemic resistance (ISR). Commonly, plant response is triggered by pathogenic microorganisms associated with the aerial portions of the plant [148]. However, also a blend of volatile organic compounds (VOCs), especially 3-hydroxy-2-butanone (acetoin) and 2,3-butanediol, emitted by the nonpathogenic rhizobacteria *B. subtilis* and *B. amyloliquefaciens,* were shown to trigger enhanced plant growth [149]. Acetolactate synthase, AlsS, catalyzes the condensation of two pyruvate molecules into acetolactate, which is decarboxylated by AlsD to acetoin. FZB42 harbors these genes that catalyze the 2,3-butanediol pathway.

(5) ***Rhizosphere competence:*** the ability of FZB42 to colonize on plant roots is a prerequisite for stimulating plant growth. The capabilities to surface adhesion, biofilms formation and swarming are essential for rhizosphere competence. Notably, three unique genes, RBAM00750, 751 and 754, encoding proteins with a collagen-related GXT structural motif (CSM) and likely to be involved in surface adhesion or biofilm formation [150] were detected in the FZB42 genome. The complete set of genes previously identified to be involved in biofilm and fruiting body formation in *B. subtili* [151-153] have been also found in FZB42. Exopolysaccharides (EPSs) participate in biofilm formation through holding the chains of cells together in bundles. The 15-gene exopolysaccharide operon *eps*A-O, which is believed to produce an EPS essentially for forming the biofilm in *B. subtilis*, exists also in the genome

of FZB42 [151-153]. The counterparts of the genes essential for swarming in *B. subtilis*, like swrA, swrB, efp [154], are present in the FZB42 genome. Our genomic analysis elucidates possible ways used by Gram-positive bacteria to colonize on plant roots [Lit. 6].

4.2 Competence for DNA uptake switches on in FZB42 during mid-exponential phase

Genetic competence is one of the survival strategies used by bacteria to adapt to certain environmental conditions. The rate at which species are able to express genetic competence in response to such conditions is likely to influence their ability to compete within their environment [155, 156]. We have developed a method for DNA transformation based on natural competence of FZB42 [18]. The results described in section 3.3.5 demonstrate that genetic competence is expressed earlier in B. *amyloliquefaciens* FZB42, than in B. *subtilis* 168. The genomic data offer several plausible explanations for this difference.

All of the genes that are involved in the genetic competence of BS168 are also present in the FZB42 genome, and most of them share 50-98% similarities in amino acid levels. Exceptions include ComGD, ComGE and ComGF, which constitute part of the DNA transport machinery and share 47%, 42%, and 46% amino acid similarities between the two genomes. Differences between the genes in FZB42 and BS168 may affect the efficiency with which they are able to bind and to take up exogenous DNA [157]. These differences may explain why the transformation rate of BS 168 was higher than that of FZB42 (3.3.5).

The most striking difference between FZB42 and BS168 in the mechanism of genetic competence is due to the quorum-sensing system. It is well accepted that accumulation of pheromones in the growth medium signals the presence of a sufficient number of congeners (a quorum), and triggers various cell-density-dependent processes [155]. ComX is the pheromone whose tryptophan residue is modified with an isoprenyl by ComQ [156]. ComX of FZB42 shows only 23% similarity to the BS168 gene, and ComQ shares only similarity of 35%. Interestingly, the ComP proteins share 52 % similarity, but their variable N-termini (1-157 aa) share only 32% similarity at the amino acid level. The N-terminus of ComP contains a binding site for isoprenyl modified ComX. These findings suggest that the altered quorum-sensing process in FZB42 may be involved in early activation of the competent pathway during mid-exponential phase.

Restriction-modification system may participate in degradation of improperly modified exogenous DNA [158]. FZB42 encodes a unique gene, RBAM_007140, which contains a conserved domain related to type I restriction-modification proteins. Curiously, the GC-content for these loci (33%) is substantially lower than the overall genome average (46.6%),

which suggests that it is the result of a gene acquisition event. RBAM_007140 can degrade presumably unmethylated DNA, as exogenous DNA. This could be another reason for the low efficiency of transformation in FZB42. This hypothesis should be tested by preparing a mutant strain deficient in RBAM_007140.

4.3 Production of secondary metabolites --- powerful tools for survival

Secondary metabolites produced by FZB42 might improve growth of FZB42 under special conditions e.g. within plant rhizosphere. In this environment, many different microorganisms compete for their survival. To deal with this challenge, FZB42 has developed several strategies to suppress growth of competitors. Producing diverse antibiotics is one possibility. More than 14% of the FZB42 genome is devoted to production of antibiotics, a figure which widely exceeds the average number of antibiotics produced by *Bacilli* (4~5%) [28]. Two main groups of antibiotics play important roles in suppressing competitors: the non-ribosomally synthesized lipopeptides, like bacillomycin D and fengycin, which are responsible to inhibit growth of fungi and the polyketides bacillaene, difficidin, and macrolactin are major secondary metabolites acting against bacterial competitors (see 4.3.1). In addition the dipeptide bacilysin suppress bacteria and fungi as well. The unknown antibiotic possibly produced by RBAM_029230 seems to provide an antibiotic with a narrow host spectrum directed mainly against its closest relatives (see 4.3.3). Another possibility to suppress indirectly growth of competitors is by depriving essential nutrients. The best studied example is production of siderophores [25, 159, 160]. The low soluble Fe^{3+} ions are abundant in soil, but are inaccessible for living organisms [160]. Siderophores, like bacillibactin produced by FZB42, make accessible these iron ions by forming soluble Fe^{3+}-siderophore complexes that are actively transported into the cell by a specific transport machinery [144].

4.3.1 Polyketides produced by B. amyloliquefaciens FZB42 are efficient in suppression of competitive bacteria

During functional analysis of the lipopeptides produced by FZB42, it rules out that wild type FZB42 is able to inhibit growth of both, fungi (such as *Fusarium oxysporum*, *Alternaria alternatae, Gaeumannomyces graminis* and *Rhizoctonia solani*) and bacteria, like *Bacillus megaterium* [8]. However, mutants impaired in lipopeptide production only lost their ability to suppress growth of fungi, but were not altered in their ability to suppress growth of bacteria [8]. This indicates presence of other antibiotics in FZB42 able to inhibit growth of competing bacteria. According to the genome sequence of FZB42, three PKS gene clusters were identified. Using single and double deletion mutants of these gene clusters made it possible to assign them as bacillaene, macrolactin and difficidin. The three polyketides showed different

levels in suppressing G ram-negative and Gram- positive bacteria, but ac tivities against fungi were not observed. Only the lipopeptide surf actin showed a very weak activity agains t bacteria ([1 19]). This result indicates that lipopeptides and polyketides produced by *B. amyloliquefaciens* FZB42 have dif ferent specificities in attacking pathogenic bacteria and fungi. Dif ferent responsibilities by these an tibiotics m ake FZB42 more ef ficient against diverse co mpetitors. Depending on presen ce of fungi or bacteria in its enviro nment, *B. amyloliquefaciens* FZB42 m ight activate dif ferent signal transduction pathways, which could switch the expression between NRPSs and PKSs.

4.3.2 nrs might be a novel siderophore

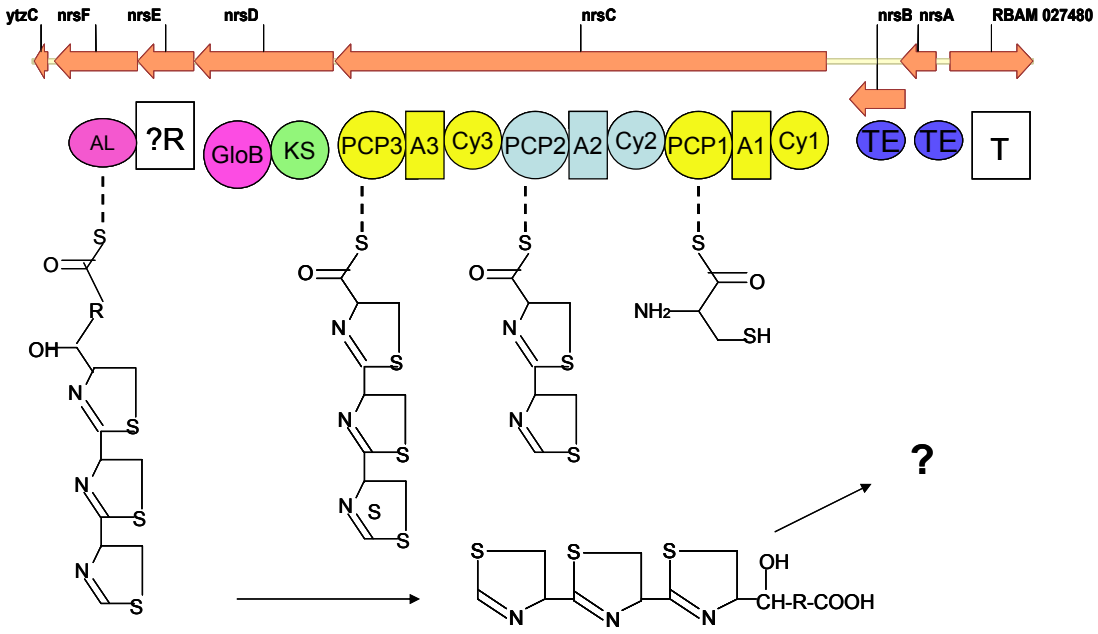


Figure 36. Hypothetical biosynthesis of nrs: A= adenylatio n dom ain A, AL = acyl Co-A synthetase, ArCP = aryl carrier protein, Cy = cy clization, PCP = peptidyl carrier protein, KS = β-ketoacyl synthase, AT = acyltransferase, M = methyltransferase, KR = ketoreductase, ACP = acyl carrier protein, O =Oxidoreductase, R = transcriptional regulator, T = transporter.

The gene cluster or ganization of *nrs* is sim ilar to that in siderophore yersiniabactin (**Figure 31**). The three condensati on dom ains in NRPS of *nrs* gene cluster m ost probably direct a cyclization activity resulting in forming a thia zol ring within the com pound, which is sim ilar to biosynthesis of yersiniabactin (**Figure 36**). T hree thiazol rings are linked directly in nrs according to the co-linear rule of PKS/NRPS biosynthesis (**Figure 36**), whereas one acyl keto-group is inserted between the s econd and the third thiazol ring by yersiniabactin synthetase. Unlike yersiniabactin synthetase, ArCP domain or other possible ACP domains are missing in front of NrsC, suggesting that the star ter unit of nrs m ight be thiazol ring (**Figure 36**). A polyketide synthase containing β-ketoacyl synthase, Zn-dependent hydrolase (GloB) and acyl

Co-A synthetase domains might link one acyl-CoA substrate or its derivatives on the tri-thiazol ring structure, but the exact substrate is unknown (**Figure 36**). The hybrid NRSP/PKS compound might be released from the last PKS module by a thioesterase (**Figure 36**). However, it is unknown whether the active form of nrs is a monomer or a multimer like bacillibactin.

Mutant strains CH39 (*dhb*$^-$) and CH40 (*nrs*$^-$ and *dhb*$^-$) did not show different growth under iron deprivation (**Figure 32**), but by CAS-assay we registered decrease of the siderophore level in the double mutant CH40 in comparison with CH39 (**Figure 33**). The amount of the siderophores in CH40 was similar to that of CH3 (*sfp*$^-$). These results suggest that nrs may be produced in FZB42 as a new siderophore. However, nrs could not be detected by HPLC, LC-ESI MS and MALDI TOF MS in the supernatant of Landy medium. This might be due to the following reasons: 1) besides bacillibactin nrs is an additional siderophore that can be synthesized only in very low amounts and in complete absence of iron. 2) Further unknown modifications occur during biosynthesis of nrs, therefore its structure and molecular weight might deviate from the structure we expected in accordance to the co-linear rule of NRPS/PKS. It has been reported that siderophores could be a subunit of bacteriocins. Microcin E492, a bacteriocin produced by enterobacteria, needs a post-translational modification to conjugate a siderophore with its peptide (**Figure 37**) [161]. It is also possible that the main function of nrs is to modify an unknown bacteriocin because of its very weak siderophore activity.

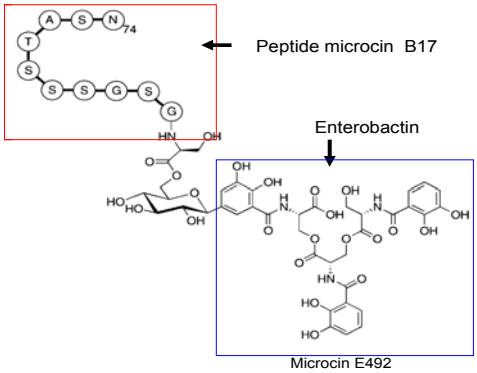


Figure 37 Structure of microcin E492: conjugation of a ribosomal peptide and non-ribosomal enterobactin

4.3.3 A new family of antibiotics in Bacilli

A unique gene cluster 3.2.2.1 in FZB42 is related to the activity of an unknown ribosomally synthesized compound that was detected during screening of sigW-dependent resistance genes [125]. Complete deletion of this gene cluster in mutant CH5 results in losing this antibiotic

activity. According to our protein sequence analysis, RBAM_029230 could be the synthesis gene for this peptide antibiotic. Cysteine residues are essential for thioester linkage with dehydrated Thr/Ser residues in post-translational modification of lantibiotics [33], but there is no cysteine in RBAM_029230 suggesting that this novel antibiotic is not a lantibiotic. According to the size of the hypothetical mature peptide (36 aa and 3455.8601 Da), this could be the first bacteriocin type II in *Bacilli*. The gene cluster related to the unknown antibiotic activity in WY01 is similar to a gene cluster occurring in *Paenibacillus larvae subsp. larvae* BRL-230010 suggesting that production of this antibiotic might be not a rare phenomenon in nature.

A strong antibiotic activity of this novel antibiotic was detected in supernatant of mutant CH5 (*sfp*$^-$ and *yczE*$^-$). However, we were unable to detect specific peaks in both the supernatant and cell surface by LC-ESI MS and MALDI-TOF MS. One of the explanations can be limitation of these methods. The peptide backbone of the novel antibiotic does not contain Trp, Tyr and Cys, so it cannot be made visible by UV spectrophotometry. Obviously, HPLC, utilizing only this principle, could not detect this novel antibiotic. The low concentration of the compound can be the second reason for our difficulties in detecting the compound. Bacteriocins show strong antibiotic activity already in a concentration less than 10 ng/ml [28]. Moreover, modifications on the backbone are common in bacteriocins [43], so the novel antibiotic could have been modified post-translationally and its molecular mass might differ strongly to the expected one. The forth possibility is that the gene cluster involving RBAM_029230 does not contain the structural gene for this unknown antibiotic and might only be involved in its transport and/or immunity.

4.4 Polyketides and their biosynthesis

The polyketide synthases of bae, mln and dfn show some striking features in the ir domain organization: 1) split module in two proteins, 2) tandemly repeated ACP domains and 3) *trans*-action of the domains.

4.4.1 Split module on two proteins

Unlike erythromycin synthases where the modules are always organized together in one protein [162], the modules of bae, mln and dfn were split almost on two different proteins. Split modules occur more usually after a KS domain. Another notable feature is the presence of the non-canonical bi-modular sequence KS KR ACP KS^0 DH ACP with a split downstream module in two proteins [78]. This unusual pattern was found twice in *bae* and *dfn* gene cluster. KS^0 is a non-elongating KS, as evidenced by the finding that the DH in the

downstream module is coupled with the KR from the upstream module, which directs formation of the double bond in the polyketide.

4.4.2 Tandemly repeated ACP domains

Duplications of ACPs within the modules of polyketide synthase type I are very common, and eight of such modules were found in FZB42. The function of these tandemly repeated ACPs is unknown. According to the activity of the repeated ACPs and the modification domains provided in the same modules, these ACPs can be classified in five groups. The group 1 contains two ACPs in a module that does not contain any modification domain, like module 14 of DfnJ (Table **13**). It is hard to assign the functions of these both ACPS, because these two ACPs provide a complete conserved sequence with a serine active site, indicating that both of them should be active. It could be speculated that one of them could be sufficient for the elongation step, but presence of two ACPs might speed up elongation. The module six of BaeL belongs to the group 2. In comparison to group 1, the sequence from one of the group 2 ACPs is not complete: ACP6-1 of BaeL possesses the serine site for attaching the Ppant moiety, but 27 amino acids of the N-terminus in ACP6-1 are missing. On the other hand, ACP6-2 contains a complete conserved sequence with a serine active site (Table **13**). It can be speculated that the speed of modifying the serine residue of the short ACP6-1 by Sfp and/or of attaching of acyl-CoA substrates might be reduced compared to ACP6-2. The first ACP of the group 3 contains the whole sequence with conserved domains except that the serine residue for modification with Ppant is missing, so only the second ACP is functional in chain elongation (Table **13**). All modules belonging to group 3 contain keto-reductase, so the first dysfunctional ACP could probably serve as barrier to separate spatio temporally the modification reaction from the Claisen condensation. This group of ACPs seems to be more rate limited. Group 4 provide two completely functional ACPs that follow the MT domain directly. This type of duplicated ACP domains was already well investigated in one of the modules from leinamycin synthase. The first ACP domain is involved in ensuing C-methylation process and the second ACP domain is used for the initial chain elongation step [163]. A unique module of Dfn synthase is module 6 that has three tandem ACPs. These three ACPs are a combination of group 3 and group 4 (Table **13**). The difference of group 4 and group 5 is that the latter has only DH and KR, but not MT domains (Table **13**). Like group 4, the first ACP can be involved in the modification process. To prove these speculations, further experiments like in-frame deletion of ACP and point mutation to substitute the serine residue by other amino acids should be performed in future.

Table **13**. Classification of the duplicated ACPs in FZB42:

Proteins	Organization of the domains	Classification of the ACPS
DfnJ	KS14,ACP14,ACP14,	Group 1
BaeL	KS6, ACP6-1, ACP6-2	Group 2
MlnD/E	KS5,ACP5-1/KR5,ACP5-2	Group3
MlnG	KS9,ACP9-1,KR9, ACP9-2	Group 3
DfnF/G	KS6,ACP6-1,DH6,KR6/MT6,ACP6-2,ACP6-3	Group 3, 4
DfnD	KS1,DH1,KR1,MT1,ACP1-1,ACP1-2	Group 4
MlnB	KS1,DH1,KR1, ACP1-1,ACP1-1	Group 5
MlnE/F	KS7/DH7,ACP7-1,ACP7-2,KR7	Group 5

4.4.3 Cross-talk of trans-ATs in different gene clusters

One of the most striking features in the *pks* gene clusters of the FZB42 genome is the organization of the acyl transferases. Three ATs in front of the *bae* gene cluster, one in front of the *dfn* and *mln* gene cluster were identified in FZB42. These PKS architecture, found recently in many polyketide synthase type I system in bacteria, is characterized by a discrete, iteratively acting AT protein that loads the extender units in trans to "AT-less" multifunctional type I PKS proteins for polyketide biosynthesis [75, 163].

The AT *trans*-action phenomena were observed until now only in genomic systems harbouring just a single PKS gene cluster [163, 164]. No investigation has been performed in systems with several AT-less PKS gene clusters. *B. amyloliquefaciens* FZB42 provides an excellent possibility to examine whether the AT-domains can act *in trans* among different *pks* gene clusters.

The AT-domains are essential for their own polyketide production, because deleting any AT-domain in its gene cluster impaired production of the respective polyketide, also in case of the three trans-ATs in front of the *bae* gene cluster. As expected, MlnAT, preceding the macrolactin gene cluster, did not participate in synthesis of dfn and bae, because the amounts of these compounds in the *mlnAT* deletion mutant were found similar to that of the wild type. Surprisingly, deletion of *dfnAT* led to a reduction in the amount of mln and bae, which implied that *dfnAT* could be engaged in both biosynthesis of mln and bae, but is not essential for their production. Assigning the specific functions of the three *baeAT*s is not possible at present, due to possible polar effects of the deletions introduced in the respective genes. However, comparison of the three mutants provided some ideas about the roles of the three AT domains located in the *bae* gene cluster. The amount of mln did not change evidently in mutant CH24 compared with CH25, whose expression of BaeAT1 was not affected. It means that BaeAT2

and BaeAT3 could contribute to biosynthesis of mln, whereas BaeAT1 is not involved in production of mln. BaeAT3 is essential for production of dfn, because it could not be restored by CH26 that expressed both Bae1AT and Bae2AT.

In summary, I got first evidence that BaeAT2, BaeAT3 and DfnAT are not only essential for synthesis of their own polyketides, but also important in biosynthesis of other polyketides. Since the production of dfn needs both BaeAT3 and DfnAT, it implies that some of these AT-domains could recognize specific substrates or connect specific ACPs. In addition, the presence of AT domains from other gene clusters could improve the production of mln and bae, but was not essential. The *trans*-effect of AT domains is not restricted on their respective gene cluster but can also enhance synthesis of other compounds synthesized by gene clusters located on fair distant sites of the genome.

4.4.4 Substrate specificity of AT domains

Bae1AT has been characterized as a malonyl-CoA specific AT that accepts AcpK as a thiolation (T) domain substrate [165]. Bae3AT, MlnAT and DfnAT are hybrid enzymes bearing also a predicted 2-nitropropane dioxygenase (NPD)-like domain. It was suggested by analysis of conserved domains that Bae3AT could probably recognize malonyl-CoA, because Bae3AT contains a FabD domain directly be hind the predicted 2-nitropropane dioxygenase (NPD)-like domain [166]. Bae2AT, DfnAT and MlnAT contain an acyl transferase domain cl08282, which, unfortunately, do not provide any clue for their target substrates.

The ^{13}C-feeding experiment demonstrated that acetate/malonate is the only substrate for the starter and the extender units in biosynthesis of mln (3.3.1.4). According to the results described in 3.3.1.4, MlnAT alone is sufficient for biosynthesis of mln. Therefore, it can be concluded that MlnAT can recognize acetyl-CoA and its derivate malonyl-CoA as substrate. Because Bae2AT, Bae3AT and DfnAT could act as ATs in biosynthesis of mln, it implies that Bae2AT, Bae3AT and DfnAT can use at least acetyl-CoA and its derivative malonyl-CoA as a substrate.

4.4.5 Roles of modification enzymes in the *pks* operons

β-keto-acyl reductase (KR), hydroxyacyl dehydratase (DH) and enoyl reductase domain (ER) are the enzymes involved in different reduction reactions of β-keto groups, which result in stepwise conversion of the keto-group to a single bond. Depending on the level of reduction of the β-keto group, one can conclude which modification domains should be present in a module. Notably, not all of these modification enzymes were detected within the modules present in the *bae*, *mln* and *dfn* gene clusters. The ER could not be detected in the corresponding modules of the *bae* gene cluster, despite that the respective double bond is

reduced to a single bond. Moreover, one KR, one DH and two ER were found absent in the *dfn* gene cluster. Two DH and one ER were not present in the *mln* gene cluster. It was supposed that the function of these enzymes could be provided by other modules or other proteins with similar function *in trans* [119]. We have already demonstrated that BaeS, a cytochrome P450 following the bacillaene structure genes is directly involved in the desaturation of the double bond in bacillaene, instead of an ER domain [Lit. 3]. A cytochrome P450 was deleted in *dfnM*, and surprisingly, no mln and dfn could be detected in HPLC according to the UV-spectrum and retention time for mln and dfn. We speculate that the cytochrome P450 in *dfnM* could act *in trans* in the *mln* and dfn gene cluster to desaturate the double bond in macrolactin and difficidin. To prove this speculation, the mutant will be further investigated by LC-ESI MS. Deletion of the TE domain in the *dfn* gene (CH10) cluster implied that the TE domain could not only catalyze releasing of its product, but also takes over partially the function of the TE domain from the bacillaene polyketide synthase. A 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) homolog in DfnL and an enoyl-CoA hydratase (EH) homolog in DfnK were supposed to be involved in formation of a exomethylene group in dfn [167]. No dfn could be detected in mutants deleted in one of these genes, but production of mln and bae was not affected. This indicates that DfnK and DfnL are only involved in biosynthesis of dfn. Further analysis of these two mutants by LC-ESI MS is necessary to detect the intermediates produced by them. Information from LC-ESI MS about CH10, CH15, CH16 and CH17 will contribute to a better understanding of the modification of polyketides in discrete AT-less polyketide synthases type I.

Except reduction of the keto-group to a single bond, there are some other modifications of polyketides, like glycosylation (3.3.1.1). These particular modifications may be required for modulation of antibiotic activity. One of the most well known polyketides is macrolactin. The macrolactin ring can be modified by attachment of a malonyl- or succinyl-residue at position 7 or a glycosidation at positions 7 or 15, these modifications lead to a different activity spectrum in the members of the macrolactin family [89, 90, 168-170]. In this work, we detect four different members of macrolactin produced by FZB42 (3.3.1.1). Furthermore, we detected another two different species of bacillaenes and hydrobacillaenes according to their MS, in addition to the isoforms already determined by Butcher et al. [130]. However, the exact chemical structures of their modifications are unknown at present (3.3.1.1). The tailoring modifications are perhaps performed by external enzymes as methyl-, acyl- and glycosyltransferases, since no such functional genes or domains have been found within the *pks* gene clusters.

4.4.6 The *pksX* gene cluster is responsible for bacillaene production

Since the discovery of the *pksX* system in the genome of *B. subtilis* 168 [7], difficidin was believed as being its metabolic product, however the *pksX* gene cluster shows high similarity to the gene cluster *bae* in FZB42 that was identified as being responsible for bacillaene synthesis. Our investigation of the supernatant of *B. subtilis* OKB105 demonstrated that not only surfactin but also bacillaene is produced when a functional *sfp* gene was introduced. We have assigned for the first time the *pksX* gene cluster for production of bacillaene, which made it possible to explain the structure of bacillaene [130, 137]. A careful comparison of both gene clusters, *pksX* and *bae*, revealed, however, minor differences: 1) *pksF* (beta-ketoacyl synthase) does not exist in the genome of FZB42. The *pks*CFGHI/*acp*K subcluster was believed to be involved in the insertion of a β-methyl branch on PksL [165]. Absence of PksF in the *bae* gene cluster of FZB42 doubts this conclusion. 2) *pksA*, the first gene in the *pksX* system, stands separate around 730 kb kb upstream of the *bae* gene cluster in FZB42.

4.4.7 Bacillaene, macrolactin and difficidin are widely distributed in nature

To prove, whether production of the diverse polyketides observed in *Bacillus amyloliquefaciens* FZB42 is a unique property, another 29 *Bacilli* strains were investigated for their ability to produce bacillaene, difficidin and macrolactin by LC-ESI MS and PCR. The result of LC-ESI MS was nicely corresponding to the result obtained by PCR-amplification of the *trans*-AT domains located within the respective gene clusters. Macrolactin was only found in the group of plant associated *B. amyloliquefaciens* strains. The PCR fragment of their single AT domain had the same size as in FZB42 suggesting that the AT domain of mln is well conserved. In addition to the *B. amyloliquefaciens* strains, difficidin seemed to be produced by some *B. licheniformis* strains according to the result from LC-ESI MS. However, PCR analysis of the AT domain implied that these strains do not possess the conserved AT domain. No other data support the idea that *B. licheniformis* is capable to produce polyketides. We conclude that the result obtained from LC-ESI is misleading. The production of bacillaene seems to be more widely distributed in different species of *Bacilli* than the production of other polyketides. The three *trans*-AT domains of the *bae* cluster are varying according to our PCR analysis. In summary, our results from PCR analysis correlate nicely with the results obtained with chemical methods. This demonstrates that PCR can serve as a powerful tool for high-throughput screening of producer strains of these polyketides.

4.5 Sfp is necessary for both, polyketide and lipopeptide production

Sfp is known to activate peptidyl carrier proteins (PCPs) of nonribosomal peptide synthetases, peptidyl carrier proteins in the NRPS-PKS hybrid (hPCP) as well as acyl carrier proteins

(ACPs) of fatty acid and polyketide synthases [99]. *In vivo* investigation of the Sfp deletion mutant CH3 confirmed that Sfp converts the *apo*-form of PCP, ACP and hPCP to the *holo*-form.

Marahiel´s group performed kinetic studies with Sfp and AcpS for hPCP, ACP and PCP. The data showed that catalytic activity of Sfp for hPCP lies between that for PCP and ACP, namely 6-fold lower than for PCP and 12-fold higher than for ACP [98]. AcpS showed high catalytic activity for ACP and very weak catalytic activity for hPCP *in vitro*, whilst no catalytic activity for PCP was determined [98]. A specific feature is that the catalytic activities of Sfp and AcpS for ACP were dependent on the concentration of Apo-ACP [98]. AcpS alone could not complement production of polyketides, lipopeptides and the siderophore in CH3 *(sfp$^-$)* (3.3.3), which indicates that Sfp is the only PPant transferase involved in production of these secondary metabolites *in vivo*. Modification of ACP, PCP and hPCP by Sfp is not only essential for both polyketides and lipopeptides production, but also determining the production rate. It seems that Sfp has a preference for production of lipopeptides than for polyketides according to semi-quantitative analysis performed with MALDI TOF MS and kinetic studies [98].

4.6 YczE might be involved in forming of a megaenzyme complex during synthesis of polyketides and lipopeptides

YczE, a hypothetical membrane protein, seems to be involved in production and/or transport of polyketides and bacillomycin D. YczE does not share similarity with known transport related proteins or conserved domains. Expression of bacillomycin D was not restored by exchange the promoter by the IPTG-dependent promoter P_{spac}, which indicates that YczE does not act as a transcriptional activator in biosynthesis of bacillomycin D [Lit. 5]. One possible role of YczE can be post-translational modification of enzymes involved in synthesis and/or export of secondary metabolites. Because biosynthesis of surfactin and fengycin was not affected in *yczE* deficient mutants, YczE is not involved in modification of Sfp. Straight *et al.* demonstrated by using Gfp, Cfp, Yfp fusion proteins, that the bacillaene synthase forms a mega-complex in *B. subtilis*. This megaenzyme complex anchors at the cell membrane, whilst surfactin synthetase was soluble in cytoplasm [171]. This suggests that the assembly line of some lipopeptides and polyketides is associated with a membrane sub-domain, either directly or by interaction with a membrane protein [171]. I hypothesize that YczE could be a candidate for a membrane protein associated with PKS or NRPS to form a megaenzyme complex essential for biosynthesis of some polyketides and lipopeptides. However, my previous

protein structure analysis provided no clue about possible binding sites of YczE with the member of megasynthase complex.

It is believed that checkpoints for transcriptional and/or posttranslational switching of the production of polyketides and lipopeptides exist in *Bacilli*. On transcriptional level, a DegU knockout mutant (TF1) has completely lost its ability to produce bacillomycin D, and CH23 (*comA*$^-$) displayed reduced production of surfactin. Another possibility is that FZB42 uses different signal transduction pathways to get information from the environment for deciding whether it should produce more antifungal lipopeptides or more polyketides acting against bacterial competitors. In this context the role of transcription regulator pksA remains also to be elucidated. In post-translational regulation, there are two candidates, Sfp and YczE, which could form a second regulatory level in governing polyketide and lipopeptide production.

4.7 Resistance against antibiotics in Bacillus amyloliquefaciens FZB42

B. amyloliquefaciens FZB42 produces at least eight antibiotics or toxins to suppress growth of competitors in its daily struggle for available ecological niches. At the same time, FZB42 must provide different mechanisms to protect themselves from antibiotics and toxins that are produced not only by themselves but also by other microorganisms.

Immunity is the mechanism of self-protection from bacteriocins. It is especially important if the bacteriocin does not need a specific receptor for its action, as in the case for the type A lantibiotics that form pores in the cytoplasmic membrane. Remarkably, the synthesis genes for the biosynthesis of known lantibiotics in *Bacilli* are absent in the genome of FZB42, but their immunity genes and regulators are present (3.2.2.1). The two-component system MrsK2/MrsR2 or SpaK/SpaR can activate the production of the immunity genes MrsEFG or SpaEFG in FZB42, if mersacidin or subtilin is produced by other competing strains in the environment. Interestingly, another plant associated *B. amyloliquefaciens* strain, B946, was shown to actively produce mersacidin (Vater unpublished observation), suggesting that this lantibiotic might be common in this environment. The immunity genes *alb*BCD are not present in FZB42, but AlbG (a membrane protein) is present in the genome of FZB42. It indicates that AlbG is probably involved in export of subtilosin A. Unlike specific immunity, the *yqeZyqfAB* gene cluster provides an unspecific antibiotic resistance in a SigW-dependent manner. It was reported that the presence of this gene cluster can protect the strains against sublancin [125]. Screening the genome of FZB42 revealed, that all genes related to sublancin synthesis are absent in FZB42. This is due to the deletion of the *spβ*2 prophage, whereas the *yqeZyqfAB* gene cluster is present. Although FZB42 does not produce any known bacteriocins, it contains a wide spectrum of resistance systems acting against different

bacteriocins. These capabilities enable the bacterium to survive within rhizosphere in permanent competition with other rhizobacteria.

Bacitracin is a cyclic dodecylpeptide antibiotic that is non-ribosomally synthesized by *B. licheniformis* and *B. subtilis* [172]. The complexity of its resistance mechanisms has been comprehensively analyzed: 1) expression of bacitracin-specific ABC transporters [173, 174], 2) *de novo* synthesis of undecaprenyl monophosphate [175], 3) expression of alternative undecaprenyl pyrophosphate phosphatases [176] and 4) exopolysaccharide production [177]. Of these possibilities, induction of bacitracin-specific ABC transporters is the most efficient resistance mechanism. FZB42 provides a bacitracin-specific ABC transporter BceAB. BceRS, a two-component system, which is located upstream of the operon and probably regulates the expression of BceAB.

The *ydbST* and *fosB* genes in a *Bacillus* strain [125] were identified as resistance genes acting against an unknown antibiotic that is ribosomally synthesized by *B. amyloliquefaciens* FZB42 (3.2.2.1). Homologues of *ydbST* and *fosB* exist also in the genome of FZB42. The *ydbST* genes encode related membrane proteins of unknown function containing one and three conserved bacterial membrane flanked domains (pfam03703).

Except surfactin, no evidence for resistance mechanism encoded in lipopeptide and polyketide gene clusters were detected in FZB42. SwrC, previously designated as multidrug efflux pump YerP, is self-resistant against surfactin [154] and is also essential for swarming motility [178]. During screening of drug efflux proteins, four transporters (YkkCD, ErbAB, YvaED and YvdSR) belonging to SMR family were identified. Site-directed mutagenesis indicated that the conserved membrane-embedded glutamates in the first transmembrane helices of both EbrA and EbrB are required for multidrug efflux activity. However, the two glutamates are non-equivalent since EbrA E15 is required for substrate binding while EbrB E14 is not [133]. Absence of glutamate residues in YvaD, the analogue of EbrA, might hinder this SMR system to bind with its substrates. The other possibility is that YvaD has a substrate spectrum, completely different to that of the known SMR members. It is reported that EbrAB and YkkCD are broad-specific multidrug efflux pumps [134, 179]. However, whether these SMR members could pump out lipopeptides or polyketides is still unknown. Cloning of these genes in *E. coli* K12, an indicator strain sensitive for FZB42 antibiotics, can be a useful approach for investigating this question.

4.8 Perspective

Availability of the whole genome sequence and its demonstrated genetic amenability allows use of *B. subtilis* 168 as model strain for investigation of molecular biology and cell biology

of Gram-positive bacteria. However, *B. subtilis* 168 is an early domesticated laboratory strain underlying several rounds of X-ray irradiation, resulting in deficiency in many features e.g. biofilm formation and nonribosomal production of several secondary metabolites. One of the well-known examples is that silencing of a functional Sfp leads to inability to produce the non-ribosomal synthesized antibiotics (bacillaene, surfactin and fengycin) and siderophore (bacillibactin) [6]. However, natural isolates of *Bacillus subtilis,* which are reflecting a metabolism unimpaired by secondary genomic mutation are often not DNA transformable under laboratory conditions. Genetic amenability, a rare feature in nature isolates, together with its ability to produce a vast array of different secondary metabolites, is one of the advantages in applying the whole sequenced FZB42 as a model strain for research, especially in context of plant growth promotion and biocontrol. Genetic manipulation of the unique genes related to plant growth promotion in FZB42, transcriptome analysis and proteome investigation will provide a deeper insight about the genes involved in such processes.
Antibiotic production is not only an interesting subject for biocontrol research but also for several applications in medicine. Dramatic increase of multidrug resistances is a big challenge for developing new antibiotics dealing with multi-resistant bacteria. The design of such drugs would be greatly facilitated by a deeper knowledge of gene organization and the structures of the underlying transport systems, which are recently poorly understood. The capacity of the antibiotics biosynthesis machinery in FZB42 is comparable with that of the well-known antibiotic producer *Streptomyces* strains, paving the way for development of novel drugs useful as biopesticides in agriculture (biocontrol) and in medicine as well. Especially biopesticides originating from FZB42 and other related plant-associated *B. amyloliquefaciens* strains will fairly enhance applying of biological tools in an efficient and more sustainable agriculture. Occurrence of AT-less PKSs in FZB42 demonstrates that the principles underlying biosynthesis of polyketides are more diverse than originally expected. Full understanding of the different underlying biological mechanisms will significantly enhance future bioengineering of tailor-made antibiotics.

References

1. Sonenshein, A.L., *Control of sporulation in itiation in Bacillus subtilis.* Curr Opin Microbiol, 2000. **3**(6): p. 561-6.
2. Paidhungat M, S.P., *Spore germination an d o utgrowth.* Bacillus subtilis and its closest Relatives: From Genes to Cells. Edited by Hoch JA, Losick R, So nenshein AL. W ashington, DC: American Society for Microbiology, 2002. : p. 537-548.
3. FG , P., *Systematics and Ecology of Bacillus.* In: Sonenshein AL, Hoch JA, Losick R (eds) Bacillus subtilis and ot her gram posit ive bact eria. Am erican Society f or Microbiology, Washington D.C. , 1993: p. 3-16.
4. Priest, F.G., Goodfellow, M., and Todd, C., *A numerical classifica tion of the ge nus Bacillus.* J Gen Microbiol, 1988. **134**(7): p. 1847-82.
5. Burkholder PR, G.N., *Induced biochemical mutations in B acillus subtilis.* Am J Bot, 1947. **33**: p. 345–348.
6. Hoch, J.A., Losick, R., and Sonenshein, A.E., *Bacillus subtilis and its closest Relatives: Fr om Genes to Cells.* American Society for Microbiology, Washington, DC, 2002.
7. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P ., Bolotin, A., Borchert, S ., Borriss, R ., Boursier , L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S ., Bruschi, C.V., Caldwell, B., Capuano, V ., Carter, N.M., Choi, S.K., Codani, J.J., Connerton , I.F., Danchin, A., and et al., *The co mplete genome sequence of the gram-positive bacterium Bacillus subtilis.* Nature, 1997. **390**(6657): p. 249-56.
8. Koumoutsi, A., Chen, X.H., Henne , A., Liesegang, H., Hitzeroth, G., Franke, P., Vater, J., and Borriss, R., *Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopepti des in Bacillus amyloliquefaciens strain FZB42.* J Bacteriol, 2004. **186**(4): p. 1084-96.
9. Krebs, B., B. Höding, S. M. Kübart, A. W orkie, H. Junge, G. Sch miedeknecht, P. Grosch, H . Bochow, and M. Heves. , *Use of Bacillus subtilis as bi ocontrol agent. 1. Activities and characterization of Bacillus subtilis strains.* J. Plant Dis. Prot. , 1998. **105**: p. 181-197.
10. Kloepper, J. W., Schroth, M. N. , *Plant growth-promoting r hizobacteria on rad ishes.* Proc. of the 4th Internat. Conf. on Plant Pathogenic Bacter . Vol. 2, S tation de Patholo gie Vegetale et Phytobacteriologie, INRA, Angers, France., 1978: p. 879-882
11. Kloepper , J.W., *Plant growth-promoting r hizobacteria as bi ological co ntrol a gents.* Soil Microbial Ecolog y: Appli cations in A gricultural and En vironmental Managem ent. F . B. Metting, Jr., ed. Marcel Dekker Inc., New York, USA., 1993: p. 255-274.
12. Lugtenberg, B.J., Dekkers, L., and Bloem berg, G.V., *Molecular determinants of r hizosphere colonization by Pseudomonas.* Annu Rev Phytopathol, 2001. **39**: p. 461-90.
13. Paulsen, I.T., Press, C.M., Ravel, J., Koba yashi, D.Y., My ers, G.S., Mavrodi, D.V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., Do dson, R.J., Du rkin, A.S., Brinkac, L.M.,

Daugherty, S .C., Sulli van, S.A., Rosov itz, M.J., G winn, M.L., Zhou, L., Sc hneider, D.J., Cartinhour, S.W., Nelson, W.C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E.A., Pierson, L.S., 3rd, Thomashow, L.S., and Loper, J.E., *Complete genome sequence of the plant commensal Pseudomonas fluorescens Pf-5.* Nat Biotechnol, 2005. **23**(7): p. 873-8.

14. Lucy, M., R eed, E., and Glick, B.R., *Applications of fr ee living pl ant gr owth-promoting rhizobacteria.* Antonie Van Leeuwenhoek, 2004. **86**(1): p. 1-25.
15. Glick, B.R., *The enhance ment of pl ant gr owth by f ree-living ba cteria.* Can. J. Microbiol. , 1995. **41**: p. 109-117.
16. Idriss, E.E., Makarewicz, O., Farouk, A., Rosner, K., Greiner, R., Bochow, H., Richter, T., and Borriss, R., *Extracellular phytase activity of Bacillu s amyloliquefaciens FZB45 contributes to its plant-growth-promoting effect.* Microbiology, 2002. **148**(Pt 7): p. 2097-109.
17. Patten, C.L. and Glick, B. R., *Bacterial biosynthesis of indole-3-acetic acid.* Can J Microbi ol, 1996. **42**(3): p. 207-20.
18. Idris, E.E., Iglesias, D.J., T alon, M., and Borriss, R., *Tryptophan-dependent pr oduction of indole-3-acetic acid (I AA) af fects level of plant gr owth pr omotion by Bacillus amyloliquefaciens FZB42.* Mol Plant Microbe Interact, 2007. **20**(6): p. 619-26.
19. Principe, A., Alvarez, F ., Castro, M.G ., Zachi, L., Fischer , S.E., Mori, G .B., and Jofre, E., *Biocontrol a nd PGPR featur es in native strains isolated from saline soils of Argentina.* Curr Microbiol, 2007. **55**(4): p. 314-22.
20. Singh, U.P., Sarma, B.K., and Singh, D.P., *Effect of plant growth-promoting Rhizobacteria and culture filtrate of Scler otium rolfsii on phenolic and salicylic acid contents in c hickpea (Cicer arietinum).* Curr Microbiol, 2003. **46**(2): p. 131-40.
21. Kamilova, F ., Lamers, G ., and Lu gtenberg, B., *Biocontrol strain Pseudomonas fluor escens WCS365 inhi bits germinat ion of Fusari um oxys porum spor es in tomato r oot exudate as we ll as subsequent formation of new spores.* Environ Microbiol, 2008.
22. Cazorla, F .M., Duckett, S.B., Ber gstrom, E.T ., Noreen, S., O dijk, R., Lugtenber g, B. J., Thomas-Oates, J.E., and Bloem berg, G.V., *Biocontrol of avocado dematoph ora r oot r ot by antagonistic Pseudomonas fluor escens PCL1606 co rrelates with the pr oduction of 2-hexyl 5-propyl resorcinol.* Mol Plant Microbe Interact, 2006. **19**(4): p. 418-28.
23. Bloemberg, G .V. and Lu gtenberg, B. J., *Molecular basis of p lant growth promotion and biocontrol by rhizobacteria.* Curr Opin Plant Biol, 2001. **4**(4): p. 343-50.
24. Kumar, V ., Kum ar, A ., and Kharwar , R.N., *Antagonistic potential of fl uorescent pseudomonads and control of charcoal rot of chickpea caused by Macrophomina phaseolina.* J Environ Biol, 2007. **28**(1): p. 15-20.
25. Boruah, H.P . and Kum ar, B.S., *Plant disease suppr ession and gr owth pr omotion by a fluorescent Pseudomonas strain.* Folia Microbiol (Praha), 2002. **47**(2): p. 137-43.

26. Chiew, Y.F., *Vancomycin-resistant enterococci.* Ann Acad Med Singapore, 1997. **26**(6): p. 808-14.

27. Rybak, M.J. and LaPlante, K.L., *Community-associated methicillin-resistant Staphylococcus aureus: a review.* Pharmacotherapy, 2005. **25**(1): p. 74-85.

28. Stein, T., *Bacillus subtilis antibiotics: structures, syntheses and specific functions.* Mol Microbiol, 2005. **56**(4): p. 845-57.

29. Jack, R.W., Tagg, J.R., and Ray, B., *Bacteriocins of gram-positive bacteria.* Microbiol Rev, 1995. **59**(2): p. 171-200.

30. Klaenhammer, T.R., *Genetics of bacteriocins produced by lactic acid bacteria.* FEMS Microbiol Rev, 1993. **12**(1-3): p. 39-85.

31. Hamon, Y. and Peron, Y., *[a Few Remarks on the Bacteriocins Produced by Grampositive Microbes.].* C R Hebd Seances Acad Sci, 1963. **257**: p. 1191-3.

32. Schnell, N., Entian, K.D., Schneider, U., Gotz, F., Zahner, H., Kellner, R., and Jung, G., *Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings.* Nature, 1988. **333**(6170): p. 276-8.

33. Sahl, H.G., Jack, R.W., and Bierbaum, G., *Biosynthesis and biological activities of lantibiotics with unique post-translational modifications.* Eur J Biochem, 1995. **230**(3): p. 827-53.

34. Moll, G.N., Roberts, G.C., Konings, W.N., and Driessen, A.J., *Mechanism of lantibiotic-induced pore-formation.* Antonie Van Leeuwenhoek, 1996. **69**(2): p. 185-91.

35. Banerjee, S. and Hansen, J.N., *Structure and expression of a gene encoding the precursor of subtilin, a small protein antibiotic.* J Biol Chem, 1988. **263**(19): p. 9508-14.

36. Anderson, H.H., Villela, G.G., Hansen, E.L., and Reed, R.K., *Some Physical and Biologic Properties of Subtilin and Other Antibiotics.* Science, 1946. **103**(2675): p. 419-420.

37. Stein, T., Borchert, S., Conrad, B., Feesche, J., Hofemeister, B., Hofemeister, J., and Entian, K.D., *Two different lantibiotic-like peptides originate from the ericin gene cluster of Bacillus subtilis A1/3.* J Bacteriol, 2002. **184**(6): p. 1703-11.

38. Paik, S.H., Chakicherla, A., and Hansen, J.N., *Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by Bacillus subtilis 168.* J Biol Chem, 1998. **273**(36): p. 23134-42.

39. Altena, K., Guder, A., Cramer, C., and Bierbaum, G., *Biosynthesis of the lantibiotic mersacidin: organization of a type B lantibiotic gene cluster.* Appl Environ Microbiol, 2000. **66**(6): p. 2565-71.

40. Brotz, H., Bierbaum, G., Leopold, K., Reynolds, P.E., and Sahl, H.G., *The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II.* Antimicrob Agents Chemother, 1998. **42**(1): p. 154-60.

41. Brotz, H., Bierbaum, G., Reynolds, P.E., and Sahl, H.G., *The lantibiotic mersacidin inhibits peptidoglycan biosynthesis at the level of transglycosylation.* Eur J Biochem, 1997. **246**(1): p. 193-9.

42. Marx, R., Stein, T., Entian, K.D., and Glaser, S.J., *Structure of the Bacillus subtilis peptide antibiotic subtilosin A determined by 1H-NMR and matrix assisted laser desorption/ionization time-of-flight mass spectrometry.* J Protein Chem, 2001. **20**(6): p. 501-6.

43. Hechard, Y. and Sahl, H. G., *Mode of action of modified and unmodified bacteriocins from Gram-positive bacteria.* Biochimie, 2002. **84**(5-6): p. 545-57.

44. Ennahar, S., Sashihara, T., Sonomoto, K., and Ishizaki, A., *Class IIa bacteriocins: biosynthesis, structure and activity.* FEMS Microbiol Rev, 2000. **24**(1): p. 85-106.

45. Maqueda, M., Sanchez-Hidalgo, M., Fernandez, M., Montalban-Lopez, M., Valdivia, E., and Martinez-Bueno, M., *Genetic features of circular bacteriocins produced by Gram-positive bacteria.* FEMS Microbiol Rev, 2008. **32**(1): p. 2-22.

46. Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F., *An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.* Nature, 1990. **348**(6297): p. 176-8.

47. Vanittanakom, N., Loeffler, W., Koch, U., and Jung, G., *Fengycin--a novel antifungal lipopeptide antibiotic produced by Bacillus subtilis F-29-3.* J Antibiot (Tokyo), 1986. **39**(7): p. 888-901.

48. Johnson, B.A., Anker, H., and Meleney, F.L., *Bacitracin: a New Antibiotic Produced by a Member of the B. Subtilis Group.* Science, 1945. **102**(2650): p. 376-377.

49. Denny, C., Zacharias, M.E., Ruiz, A. L., do Carmo, E. d.A.M., Bittrich, V., Kohn, L.K., de Oliveira Sousa, I.M., Rodrigues, R.A., de Carvalho, J.E., and Foglio, M.A., *Antiproliferative properties of polyketides isolated from Virola sebifera leaves.* Phytother Res, 2008. **22**(1): p. 127-30.

50. Kwon, H.C., Kauffman, C.A., Jensen, P.R., and Fenical, W., *Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "marinispora".* J Am Chem Soc, 2006. **128**(5): p. 1622-32.

51. Menendez, N., Nur-e-Alam, M., Brana, A.F., Rohr, J., Salas, J.A., and Mendez, C., *Biosynthesis of the antitumor chromomycin A3 in Streptomyces griseus: analysis of the gene cluster and rational design of novel chromomycin analogs.* Chem Biol, 2004. **11**(1): p. 21-32.

52. Lee, R.H., Slate, D.L., Moretti, R., Alvi, K.A., and Crews, P., *Marine sponge polyketide inhibitors of protein tyrosine kinase.* Biochem Biophys Res Commun, 1992. **184**(2): p. 765-72.

53. Nielsen, J.B., Hsu, M.J., Byrne, K.M., and Kaplan, L., *Biosynthesis of the immunosuppressant immunomycin: the enzymology of pipecolate incorporation.* Biochemistry, 1991. **30**(23): p. 5789-96.

54. Baumler, A., Koebnik, R., Stojiljkovic, I., Heesemann, J., Braun, V., and Hantke, K., *Survey on newly characterized iron uptake systems of Yersinia enterocolitica.* Zentralbl Bakteriol, 1993. **278**(2-3): p. 416-24.

55. May, J.J., Wendrich, T.M., and Marahiel, M.A., *The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin.* J Biol Chem, 2001. **276**(10): p. 7209-17.

56. Konz, D. and Marahiel, M.A., *How do peptide synthetases generate structural diversity?* Chem Biol, 1999. **6**(2): p. R39-48.

57. Patel, H.M., Tao, J., and Walsh, C.T., *Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from Pseudomonas aeruginosa.* Biochemistry, 2003. **42**(35): p. 10514-27.

58. Stein, T., Kluge, B., Vater, J., Franke, P., Otto, A., and Wittmann-Liebold, B., *Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4'-phosphopantetheine.* Biochemistry, 1995. **34**(14): p. 4633-42.

59. Linne, U., Doekel, S., and Marahiel, M.A., *Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases.* Biochemistry, 2001. **40**(51): p. 15824-34.

60. Hillson, N.J. and Walsh, C.T., *Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: mutant domain activity regain and ultracentrifugation studies.* Biochemistry, 2003. **42**(3): p. 766-75.

61. Sieber, S.A., Linne, U., Hillson, N.J., Roche, E., Walsh, C.T., and Marahiel, M.A., *Evidence for a monomeric structure of nonribosomal Peptide synthetases.* Chem Biol, 2002. **9**(9): p. 997-1008.

62. Menkhaus, M., Ullrich, C., Kluge, B., Vater, J., Vollenbroich, D., and Kamp, R.M., *Structural and functional organization of the surfactin synthetase multienzyme system.* J Biol Chem, 1993. **268**(11): p. 7678-84.

63. Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P., and Vater, J., *Structural and functional organization of the fengycin synthetase multienzyme system from Bacillus subtilis b213 and A1/3.* Chem Biol, 1999. **6**(1): p. 31-41.

64. Tsuge, K., Akiyama, T., and Shoda, M., *Cloning, sequencing, and characterization of the iturin A operon.* J Bacteriol, 2001. **183**(21): p. 6265-73.

65. Nakano, M.M., Marahiel, M.A., and Zuber, P., *Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis.* J Bacteriol, 1988. **170**(12): p. 5662-8.

66. Moyne, A.L., Cleveland, T.E., and Tuzun, S., *Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D.* FEMS Microbiol Lett, 2004. **234**(1): p. 43-9.

67. Ishihara, H., Takoh, M., Nishibayashi, R., and Sato, A., *Distribution and variation of bacitracin synthetase gene sequences in laboratory stock strains of Bacillus licheniformis.* Curr Microbiol, 2002. **45**(1): p. 18-23.

68. Konz, D., Klens, A., Schorgendorfer, K., and Marahiel, M.A., *The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases.* Chem Biol, 1997. **4**(12): p. 927-37.

69. Storm, D.R., *Mechanism of bacitracin action: a specific lipid-peptide interaction.* Ann N Y Acad Sci, 1974. **235**(0): p. 387-98.

70. Peypoux, F., Bonmatin, J.M., Labbe, H., Grangemard, I., Das, B.C., Ptak, M., Wallach, J., and Michel, G., *[Ala4]surfactin, a novel isoform from Bacillus subtilis studied by mass and NMR spectroscopies.* Eur J Biochem, 1994. **224**(1): p. 89-96.

71. Staunton, J. and Weissman, K.J., *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**(4): p. 380-416.

72. Shen, B., *Biosynthesis of aromatic polyketides.* Curr. Top Chem., 2000. **209**: p. 1–51.

73. Moore, B.S. and Hopke, J.N., *Discovery of a new bacterial polyketide biosynthetic pathway.* Chembiochem, 2001. **2**(1): p. 35-8.

74. Hardt, I.H., Steinmetz, H., Gerth, K., Sasse, F., Reichenbach, H., and Hofle, G., *New natural epothilones from Sorangium cellulosum, strains So ce90/B2 and So ce90/D13: isolation, structure elucidation, and SAR studies.* J Nat Prod, 2001. **64**(7): p. 847-56.

75. Petkovic, H., Sandmann, A., Challis, I.R., Hecht, H.J., Silakowski, B., Low, L., Beeston, N., Kuscer, E., Garcia-Bernardo, J., Leadlay, P.F., Kendrew, S.G., Wilkinson, B., and Muller, R., *Substrate specificity of the acyl transferase domains of EpoC from the epothilone polyketide synthase.* Org Biomol Chem, 2008. **6**(3): p. 500-6.

76. Ansari, M.Z., Yadav, G., Gokhale, R.S., and Mohanty, D., *NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W405-13.

77. Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Muller, R., *New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1.* J Biol Chem, 1999. **274**(52): p. 37391-9.

78. Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., and Piel, J., *Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection.* Nat Biotechnol, 2008. **26**(2): p. 225-33.

79. Perlova, O., Gerth, K., Kaiser, O., Hans, A., and Muller, R., *Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain Sorangium cellulosum So ce56.* J Biotechnol, 2006. **121**(2): p. 174-91.

80. Piel, J., *A polyketide s ynthase-peptide synthetase ge ne cluster fr om an uncult ured bacterial symbiont of Paederus beetles.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14002-7.

81. Martel, R.R., Klicius, J., and Galet, S., *Inhibition of the immune response by rapamycin, a new antifungal antibiotic.* Can J Physiol Pharmacol, 1977. **55**(1): p. 48-51.

82. Hatanaka, H., Kino, T ., Miyata, S., Inam ura, N., Kuroda, A., Goto, T ., T anaka, H., and Okuhara, M., *FR-900520 and FR-900523, novel immunos uppressants iso lated fr om a Streptomyces. II. Fermentation, isolation and physico-chemical and biological characteristics.* J Antibiot (Tokyo), 1988. **41**(11): p. 1592-601.

83. Miyake, H., Suzuki, Y., and Kunishige, N., *[Anticancer chemotherapy apropos of bleomycin].* Rev Laryngol Otol Rhinol (Bord), 1972. **93**(11): p. 677-89.

84. Kawamata, J., Fujita, H., and Ikegam i, R., *Anticancer effect of actinomycinic acid.* J Antibiot (Tokyo), 1960. **13**: p. 415.

85. Mishima, H., Ide, J., M uramatsu, S., a nd Ono, M., *Milbemycins, a new famil y of macr olide antibiotics. Structure determination of milbemycins D, E, F, G, H, J and K.* J Antibiot (Tokyo), 1983. **36**(8): p. 980-90.

86. Patel, P.S., Huang, S., Fis her, S., Pirni k, D., Aklonis, C., Dean, L., Meyers, E., Fernandes, P., and Ma yerl, F., *Bacillaene, a novel inhibitor of procaryotic pr otein synthesis pr oduced by Bacillus subt ilis: pr oduction, t axonomy, isola tion, physico-chemical characterization and biological activity.* J Antibiot (Tokyo), 1995. **48**(9): p. 997-1003.

87. Wilson, K.E., Flor, J.E., Schwartz, R.E., Joshua, H., Smith, J.L., Pelak, B.A., Liesch, J.M., and Hensens, O. D., *Difficidin and oxydif ficidin: novel br oad spectrum antibacte rial antibi otics produced by Bacillus subti lis. II. Isolat ion and physico-chemical characterization.* J Ant ibiot (Tokyo), 1987. **40**(12): p. 1682-91.

88. Zimmerman, S.B., Schwartz, C.D., Monaghan, R.L., Pelak, B.A., Weissberger, B., Gilfillan, E.C., Mochales, S., Her nandez, S., Currie, S.A., T ejera, E., and et al., *Difficidin and oxydifficidin: novel br oad spectrum antibacterial antibiotics pr oduced by Bacillus subtilis. I. Production, taxonomy and antibacterial activity.* J Antibiot (Tokyo), 1987. **40**(12): p. 1677-81.

89. Nagao, T ., Adachi, K., Sakai, M., Nishijima, M., and Sano, H., *Novel macr olactins as antibiotic lactones from a marine bacterium.* J Antibiot (Tokyo), 2001. **54**(4): p. 333-9.

90. Jaruchoktaweechai, C., Suwanborirux, K., Tanasupawatt, S., Kittakoop, P., and Menasveta, P., *New macrolactins from a marine Bacillus sp. Sc026.* J Nat Prod, 2000. **63**(7): p. 984-6.

91. Albertini, A.M., Caram ori, T., Scoffone, F., Scotti, C ., and Galizzi, A., *Sequence around the 159 degr ee r egion of the Bacillus subtilis genome: the pksX locus spans 33.6 kb.* Microbiology, 1995. **141 (Pt 2)**: p. 299-309.

92. Scotti, C., Piatti, M., Cuzzoni, A., Perani , P ., Tognoni, A., Grandi, G ., Galizzi, A., and Albertini, A.M., *A B acillus subtilis lar ge ORF coding for a pol ypeptide highly similar to polyketide synthases.* Gene, 1993. **130**(1): p. 65-71.

93. Mootz, H.D., Schor gendorfer, K., and Marahiel, M.A., *Functional characterization of 4'-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of Saccharomyces cerevisiae lys5.* FEMS Microbiol Lett, 2002. **213**(1): p. 51-7.

94. Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M.A., Kruft, V., and Wittmann-Liebold, B., *An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase.* J Biol Chem, 1991. **266**(34): p. 23135-41.

95. Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T., *A new enzyme superfamily - the phosphopantetheinyl transferases.* Chem Biol, 1996. **3**(11): p. 923-36.

96. Gehring, A.M., Lambalot, R.H., Vogel, K.W., Drueckhammer, D.G., and Walsh, C.T., *Ability of Streptomyces spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for E. coli holo-ACP synthase.* Chem Biol, 1997. **4**(1): p. 17-24.

97. Fichtlscherer, F., Wellein, C., Mittag, M., and Schweizer, E., *A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation.* Eur J Biochem, 2000. **267**(9): p. 2666-71.

98. Mofid, M.R., Finking, R., and Marahiel, M.A., *Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in non ribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp.* J Biol Chem, 2002. **277**(19): p. 17023-31.

99. Mootz, H.D., Finking, R., and Marahiel, M.A., *4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis.* J Biol Chem, 2001. **276**(40): p. 37289-98.

100. Walker, J.E. and Abraham, E.P., *The structure of bacilysin and other products of Bacillus subtilis.* Biochem J, 1970. **118**(4): p. 563-70.

101. Steinborn, G., Hajirezaei, M.R., and Hoffmeister, J., *bac genes for recombinant bacilysin and anticapsin production in Bacillus host strains.* Arch Microbiol, 2005. **183**(2): p. 71-9.

102. Chimara, H., *Inhibition of glucosamine synthase by bacilysin and anticapsin.* J Gen Microbiol, 1985. **131**(2): p. 265-71.

103. Kunst, F. and Rapoport, G., *Salt stress is an environmental signal affecting degradative enzyme synthesis in Bacillus subtilis.* J Bacteriol, 1995. **177**(9): p. 2403-7.

104. Spizizen, J., *Transformation of Biochemically Deficient Strains of Bacillus Subtilis by Deoxyribonucleate.* Proc Natl Acad Sci U S A, 1958. **44**(10): p. 1072-8.

105. Koumoutsi, A., *Functional genome analysis of the plant-growth promoting bacterium Bacillus amyloliquefaciens strain FZB42; characterizing its production and regulation of nonribosomal peptide synthetases.* 2006.

106. Koumoutsi, A., Chen, X.H., Vater, J., and Borriss, R., *DegU and YczE positively regulate the synthesis of bacillomycin D by Bacillus amyloliquefaciens strain FZB42.* Appl Environ Microbiol, 2007. **73**(21): p. 6953-64.

107. Stragier, P., Bonamy, C., and Karmazyn-Campelli, C., *Processing of a sporulation sigma factor in Bacillus subtilis: how morphological structure could control gene expression.* Cell, 1988. **52**(5): p. 697-704.

108. Antoniewski, C., Savelli, B., and Stragier, P., *The spoIIJ gene, which regulates early developmental steps in Bacillus subtilis, belongs to a class of environmentally responsive genes.* J Bacteriol, 1990. **172**(1): p. 86-93.

109. Ceglowski, P. and Alonso, J.C., *Gene organization of the Streptococcus pyogenes plasmid pDB101: sequence analysis of the orfeta-copS region.* Gene, 1994. **145**(1): p. 33-9.

110. Steinmetz, M. and Richter, R., *Plasmids designed to alter the antibiotic resistance expressed by insertion mutations in Bacillus subtilis, through in vivo recombination.* Gene, 1994. **142**(1): p. 79-83.

111. Kaltwasser, M., Wiegert, T., and Schumann, W., *Construction and application of epitope- and green fluorescent protein-tagging integration vectors for Bacillus subtilis.* Appl Environ Microbiol, 2002. **68**(5): p. 2624-8.

112. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a laboratory manual.* Cold Spring Harbor Laboratory, NY., 1989.

113. Schneider, K., Chen, X.H., Vater, J., Franke, P., Nicholson, G., Borriss, R., and Sussmuth, R.D., *Macrolactin is the polyketide biosynthesis product of the pks2 cluster of Bacillus amyloliquefaciens FZB42.* J Nat Prod, 2007. **70**(9): p. 1417-23.

114. Dieffenbach, C.W. and Dveksler, G.S., *PCR primer, a laboratory manual.* Cold Spring Harbor Laboratory Press., 1995.

115. Burgart, L.J., Robinson, R.A., Heller, M.J., Wilke, W.W., Iakoubova, O.K., and Cheville, J.C., *Multiplex polymerase chain reaction.* Mod Pathol, 1992. **5**(3): p. 320-3.

116. Elnifro, E.M., Ashshi, A.M., Cooper, R.J., and Klapper, P.E., *Multiplex PCR: optimization and application in diagnostic virology.* Clin Microbiol Rev, 2000. **13**(4): p. 559-70.

117. Tettelin, H., Radune, D., Kasif, S., Khouri, H., and Salzberg, S.L., *Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project.* Genomics, 1999. **62**(3): p. 500-7.

118. Landy, M., Warren, G.H., Roseman, S.B., and Colio, L.G., *Bacillomycin, an antibiotic from Bacillus subtilis active against pathogenic fungi.* Proc. Soc. Exp. Biol. Med., 1948. **67**: p. 539-541.

119. Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A.W., Gottschalk, G., Sussmuth, R.D., and Borriss, R., *Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42.* J Bacteriol, 2006. **188**(11): p. 4024-36.

120. Leenders, F., Vater, J., Stein, T., and Franke, P., *Characterization of the binding site of the tripeptide intermediate D-Phenylalanyl L-prolyl-L-valine in gramicidin S biosynthesis.* J Biol Chem, 1998. **273**(29): p. 18011-4.

121. Schwyn, B. and Neilands, J.B., *Universal chemical assay for the detection and determination of siderophores.* Anal Biochem, 1987. **160**(1): p. 47-56.

122. Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N., and Caskey, C.T., *Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification.* Nucleic Acids Res, 1988. **16**(23): p. 11141-56.

123. Tech, M. and Merkl, R., *YACOP: Enhanced gene prediction obtained by a combination of existing methods.* In Silico Biol, 2003. **3**(4): p. 441-51.

124. Pohlmann, A., Fricke, W.F., Reinecke, F., Kusian, B., Liesegang, H., Cramm, R., Eitinger, T., Ewering, C., Potter, M., Schwartz, E., Strittmatter, A., Voss, I., Gottschalk, G., Steinbuchel, A., Friedrich, B., and Bowien, B., *Genome sequence of the bioplastic-producing "Knall gas" bacterium Ralstonia eutropha H16.* Nat Biotechnol, 2006. **24**(10): p. 1257-62.

125. Butcher, B.G. and Helmann, J.D., *Identification of Bacillus subtilis sigma-dependent genes that provide intrinsic resistance to antimicrobial compounds produced by Bacilli.* Mol Microbiol, 2006. **60**(3): p. 765-82.

126. Hemphill, H.E., Gage, I., Zahler, S.A., and Korman, R.Z., *Prophage-mediated production of a bacteriocinlike substance by SP beta lysogens of B acillus subtilis.* Can J Microbiol, 1980. **26**(11): p. 1328-33.

127. Dorenbos, R., Stein, T., Kabel, J., Bruand, C., Bolhuis, A., Bron, S., Quax, W.J., and Van Dijl, J.M., *Thiol-disulfide oxidoreductases are essential for the production of the lantibiotic sublancin 168.* J Biol Chem, 2002. **277**(19): p. 16682-8.

128. Qin, X., Evans, J.D., Aronstein, K.A., Murray, K.D., and Weinstock, G.M., *Genome sequences of the honey bee pathogens Paenibacillus larvae and Ascosphaera apis.* Insect Mol Biol, 2006. **15**(5): p. 715-8.

129. Hofemeister, J., Conrad, B., Adler, B., Hofemeister, B., Feesche, J., Kucheryava, N., Steinborn, G., Franke, P., Grammel, N., Zwintscher, A., Leenders, F., Hitzeroth, G., and Vater, J., *Genetic analysis of the biosynthesis of non- ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by Bacillus subtilis A1/3.* Mol Genet Genomics, 2004. **272**(4): p. 363-78.

130. Butcher, R.A., Schroeder, F.C., Fischbach, M.A., Straight, P.D., Kolter, R., Walsh, C.T., and Clardy, J., *The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1506-9.

131. Koroglu, T.E., Kurt-Gur, G., Unlu, E.C., and Yazgan-Karatas, A., *The novel gene yvfI in Bacillus subtilis is essential for bacilysin biosynthesis.* Antonie Van Leeuwenhoek, 2008.

132. Paulsen, I.T., Skurray, R.A., Tam, R., Saier, M.H., Jr., Turner, R.J., Weiner, J.H., Goldberg, E.B., and Grinius, L.L., *The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs.* Mol Microbiol, 1996. **19**(6): p. 1167-75.

133. Zhang, Z., Ma, C., Pornillos, O., Xiu, X., Chang, G., and Saier, M.H., Jr., *Functional characterization of the heteroligomeric EbrAB multidrug efflux transporter of Bacillus subtilis.* Biochemistry, 2007. **46**(17): p. 5218-25.

134. Jack, D.L., Storms, M.L., Tchieu, J.H., Paulsen, I.T., and Saier, M.H., Jr., *A broad-specificity multidrug efflux pump requiring a pair of homologous SMR-type proteins.* J Bacteriol, 2000. **182**(8): p. 2311-3.

135. Yerushalmi, H. and Schuldiner, S., *An essential glutamyl residue in EmrE, a multidrug antiporter from Escherichia coli.* J Biol Chem, 2000. **275**(8): p. 5264-9.

136. Muth, T.R. and Schuldiner, S., *A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE.* Embo J, 2000. **19**(2): p. 234-40.

137. Moldenhauer, J., Chen, X.H., Borriss, R., and Piel, J., *Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family.* Angew Chem Int Ed Engl, 2007. **46**(43): p. 8195-7.

138. Tsai, S.C., Miercke, L.J., Krucinski, J., Gokhale, R., Chen, J.C., Foster, P.G., Cane, D.E., Khosla, C., and Stroud, R.M., *Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 14808-13.

139. Reva, O.N. and Tummler, B., *Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns.* BMC Bioinformatics, 2004. **5**: p. 90.

140. Yao, S., Gao, X., Fuchsbauer, N., Hillen, W., Vater, J., and Wang, J., *Cloning, sequencing, and characterization of the genetic region relevant to biosynthesis of the lipopeptides iturin A and surfactin in Bacillus subtilis.* Curr Microbiol, 2003. **47**(4): p. 272-7.

141. Pelludat, C., Rakin, A., Jacobi, C.A., Schubert, S., and Heesemann, J., *The yersiniabactin biosynthetic gene cluster of Yersinia enterocolitica: organization and siderophore-dependent regulation.* J Bacteriol, 1998. **180**(3): p. 538-46.

142. Keating, T.A., Suo, Z., Ehmann, D.E., and Walsh, C.T., *Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain.* Biochemistry, 2000. **39**(9): p. 2297-306.

143. Makarewicz, O., Dubrac, S., Msadek, T., and Borriss, R., *Dual role of the PhoP approximately Response regulator: Bacillus amyloliquefaciens FZB45 phytase gene transcription is directed by positive and negative interactions with the phyC promoter.* J Bacteriol, 2006. **188**(19): p. 6953-65.

144. Miethke, M., Klotz, O., Linne, U., May , J.J., Beckering, C.L., a nd Marahiel, M.A., *Ferri-bacillibactin uptake and hydr olysis in Bacillus subti lis.* Mol Micr obiol, 2006. **61**(6): p. 1413-27.

145. Zimmer, W ., Aparicio, C., and Elm erich, C., *Relationship between tryptopha n biosynthesi s and indole-3-acetic acid pr oduction i n Azospirillum: identific ation and sequencing of a trpGDC cluster.* Mol Gen Genet, 1991. **229**(1): p. 41-51.

146. Basse, C.W ., Lottspeich, F ., S teglich, W ., and Kahmann, R., *Two pote ntial in dole-3-acetaldehyde dehydrogenases in the phytopathogenic fungus Ustilago maydis.* Eur J Biochem, 1996. **242**(3): p. 648-56.

147. Hillebrand, H., Bartling, D., and W eiler, E.W ., *Structural an alysis of the nit 2/nit1/nit3 gene cluster encoding nitrilases, enzym es catalyzing the terminal activation step in indole-acetic acid biosynthesis in Arabidopsis thaliana.* Plant Mol Biol, 1998. **36**(1): p. 89-99.

148. van Loon, L.C., Bakker, P.A., and Pieterse, C.M., *Systemic resistance induced by r hizosphere bacteria.* Annu Rev Phytopathol, 1998. **36**: p. 453-83.

149. Ryu, C.M., F arag, M.A., Hu, C.H., Redd y, M.S., K loepper, J.W ., and Pare, P .W., *Bacterial volatiles induce systemic resistance in Arabidopsis.* Plant Physiol, 2004. **134**(3): p. 1017-26.

150. Rasmussen, M., Jacobsson, M., and Bj orck, L., *Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins.* J Biol Chem, 2003. **278**(34): p. 32313-6.

151. Kobay ashi, K., *Bacillus subtilis pellicle formation pr oceeds thr ough gene tically defined morphological changes.* J Bacteriol, 2007. **189**(13): p. 4920-31.

152. Branda, S.S., Gonzalez-Pastor , J.E., Derv yn, E., Ehr lich, S.D., L osick, R., and Kolter , R., *Genes involved in f ormation of structur ed multicell ular co mmunities by Bacillus subtilis.* J Bacteriol, 2004. **186**(12): p. 3970-9.

153. Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R., *Fruiting body formation by Bacillus subtilis.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11621-6.

154. Tsuge, K., Ohata, Y ., and Shoda, M. , *Gene ye rP, involved in su rfactin se lf-resistance in Bacillus subtilis.* Antimicrob Agents Chemother, 2001. **45**(12): p. 3566-73.

155. Hamoen, L.W., Kausche, D., Marahiel, M.A., van Sinderen, D ., Venema, G., and Serror , P., *The Bacillus subtilis transition state regulator AbrB binds to the -35 promoter region of comK.* FEMS Microbiol Lett, 2003. **218**(2): p. 299-304.

156. Grossman, A.D., *Genetic networks contr olling the init iation of sporulation and t he development of genetic competence in Bacillus subtilis.* Annu Rev Genet, 199 5. **29**: p . 477-508.

157. Chung, Y.S., Breidt, F., and Dubnau, D., *Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of Bacill us subtilis.* Mol Microbiol, 1998. **29**(3): p. 905-13.

158. Tock, M.R. and Dryden, D.T., *The biology of restriction and anti-restriction.* Curr Opin Microbiol, 2005. **8**(4): p. 466-72.

159. Miethke, M. and Marahiel, M.A., *Siderophore-based iron acquisition and pathogen control.* Microbiol Mol Biol Rev, 2007. **71**(3): p. 413-51.

160. de Weger, L.A., van Aren donk, J.J., Recourt, K., van der Hofstad, G.A., Weisbeek, P.J., and Lugtenberg, B., *Siderophore-mediated uptake of Fe3+ by the plant growth-stimulating Pseudomonas putida strain WCS358 and by other rhizosphere microorganisms.* J Bacteriol, 1988. **170**(10): p. 4693-8.

161. Nolan, E.M., Fischbach, M.A., Koglin, A., and Walsh, C.T., *Biosynthetic tailoring of microcin E492m: post-translational modification affords an antibacterial siderophore-peptide conjugate.* J Am Chem Soc, 2007. **129**(46): p. 14336-47.

162. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L., *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-9.

163. Tang, G.L., Cheng, Y.Q., and Shen, B., *Polyketide chain skipping mechanism in the biosynthesis of the hybrid nonribosomal peptide-polyketide antitumor antibiotic leinamycin in Streptomyces atroolivaceus S-140.* J Nat Prod, 2006. **69**(3): p. 387-93.

164. Piel, J., Hui, D., Fusetani, N., and Matsunaga, S., *Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia.* Environ Microbiol, 2004. **6**(9): p. 921-7.

165. Calderone, C.T., Kowtoniuk, W.E., Kelleher, N.L., Walsh, C.T., and Dorrestein, P.C., *Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8977-82.

166. Ha, J.Y., Min, J.Y., Lee, S.K., Kim, H.S., Kim do, J., Kim, K.H., Lee, H.H., Kim, H.K., Yoon, H.J., and Suh, S.W., *Crystal structure of 2-nitropropane dioxygenase complexed with FMN and substrate. Identification of the catalytic base.* J Biol Chem, 2006. **281**(27): p. 18660-7.

167. Simunovic, V., Zapp, J., Rachid, S., Krug, D., Meiser, P., and Muller, R., *Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases.* Chembiochem, 2006. **7**(8): p. 1206-20.

168. Yoo, J.S., Zheng, C.J., Lee, S., Kwak, J.H., and Kim, W.G., *Macrolactin N, a new peptide deformylase inhibitor produced by Bacillus subtilis.* Bioorg Med Chem Lett, 2006. **16**(18): p. 4889-92.

169. Romero-Tabarez, M., Jansen, R., Sylla, M., Lunsdorf, H., Haussler, S., Santosa, D.A., Timmis, K.N., and Molinari, G., *7-O-malonyl macrolactin A, a new macrolactin antibiotic from Bacillus subtilis active against methicillin-resistant Staphylococcus aureus, vancomycin-resistant enterococci, and a small-colony variant of Burkholderia cepacia.* Antimicrob Agents Chemother, 2006. **50**(5): p. 1701-9.

170. Choi, S.W., Bai, D.H., Yu, J.H., and Shin, C.S., *Characteristics of the squalene synthase inhibitors produced by a Streptomyces species isolated from soils.* Can J Microbiol, 2003. **49**(11): p. 663-8.

171. Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., and Kolter, R., *A singular enzymatic megacomplex from Bacillus subtilis.* Proc Natl Acad Sci U S A, 2007. **104**(1): p. 305-10.

172. Lukin, A.A., Rozov, A.N., and Planutene, M.V., *[Disordered bacitracin synthesis and sporulation in mutant Bacillus licheniformis with decreased proteolytic activity].* Antibiot Med Biotekhnol, 1986. **31**(6): p. 405-8.

173. Podlesek, Z., Comino, A., Herzog-Velikonja, B., Zgur-Bertok, D., Komel, R., and Grabnar, M., *Bacillus licheniformis bacitracin-resistance ABC transporter: relationship to mammalian multidrug resistance.* Mol Microbiol, 1995. **16**(5): p. 969-76.

174. Manson, J.M., Keis, S., Smith, J.M., and Cook, G.M., *Acquired bacitracin resistance in Enterococcus faecalis is mediated by an ABC transporter and a novel regulatory protein, BcrR.* Antimicrob Agents Chemother, 2004. **48**(10): p. 3743-8.

175. Chalker, A.F., Ingraham, K.A., Lunsford, R.D., Bryant, A.P., Bryant, J., Wallis, N.G., Broskey, J.P., Pearson, S.C., and Holmes, D.J., *The bacA gene, which determines bacitracin susceptibility in Streptococcus pneumoniae and Staphylococcus aureus, is also required for virulence.* Microbiology, 2000. **146 (Pt 7)**: p. 1547-53.

176. Cao, M. and Helmann, J.D., *Regulation of the Bacillus subtilis bcrC bacitracin resistance gene by two extracytoplasmic function sigma factors.* J Bacteriol, 2002. **184**(22): p. 6123-9.

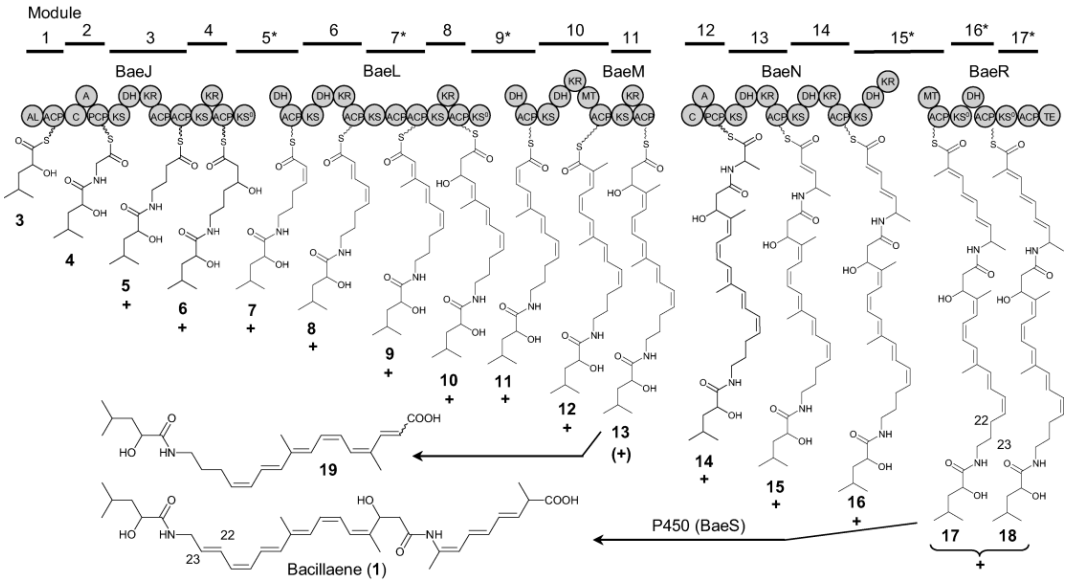
177. Tsuda, H., Yamashita, Y., Shibata, Y., Nakano, Y., and Koga, T., *Genes involved in bacitracin resistance in Streptococcus mutans.* Antimicrob Agents Chemother, 2002. **46**(12): p. 3756-64.

178. Kearns, D.B., Chu, F., Rudner, R., and Losick, R., *Genes governing swarming in Bacillus subtilis and evidence for a phase variation mechanism controlling surface motility.* Mol Microbiol, 2004. **52**(2): p. 357-69.

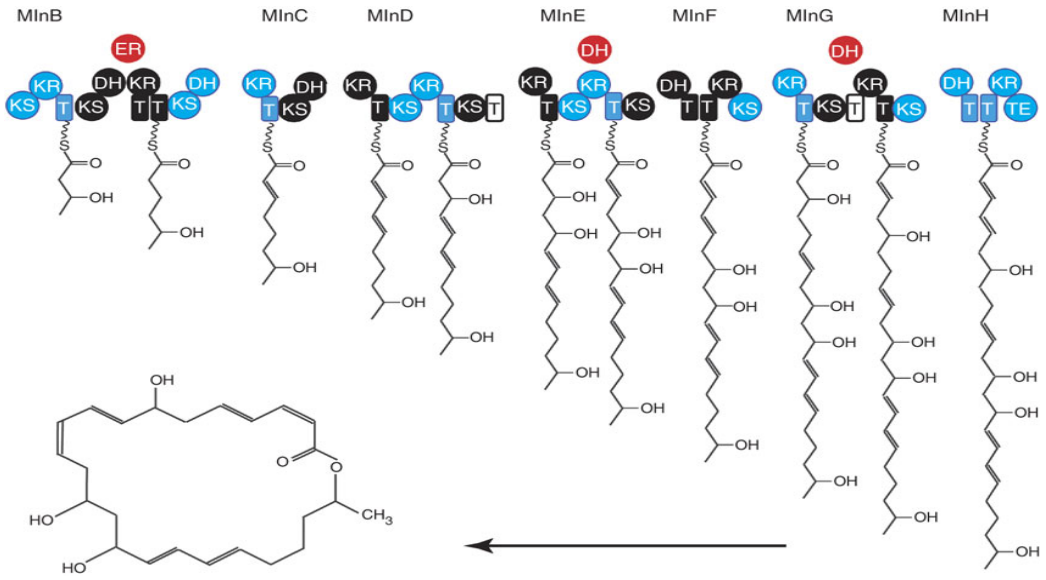
179. Kikukawa, T., Nara, T., Araiso, T., Miyauchi, S., and Kamo, N., *Two-component bacterial multidrug transporter, EbrAB: Mutations making each component solely functional.* Biochim Biophys Acta, 2006. **1758**(5): p. 673-9.

5 Appendix 1:

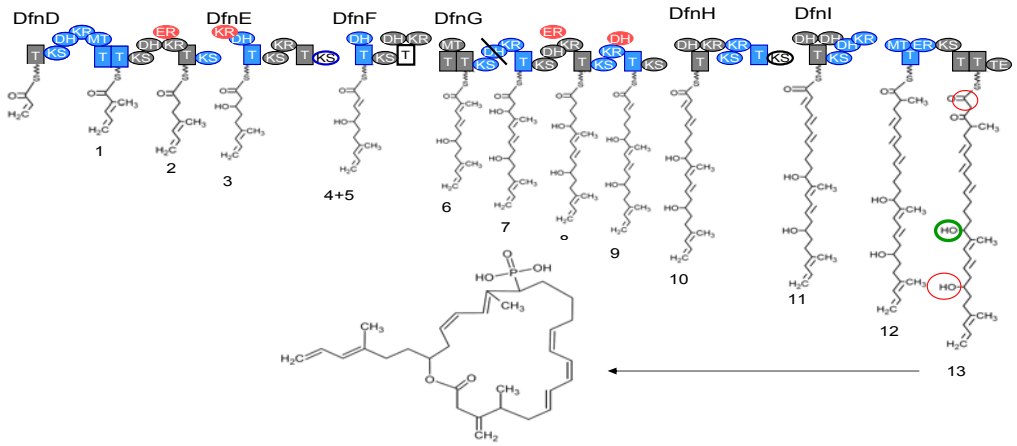
Supplementary figure **1**. bae biosynthesis pathway [**Lit. 3**]:



Supplementary figure **2**. mln biosynthesis pathway [**Lit. 4**]:

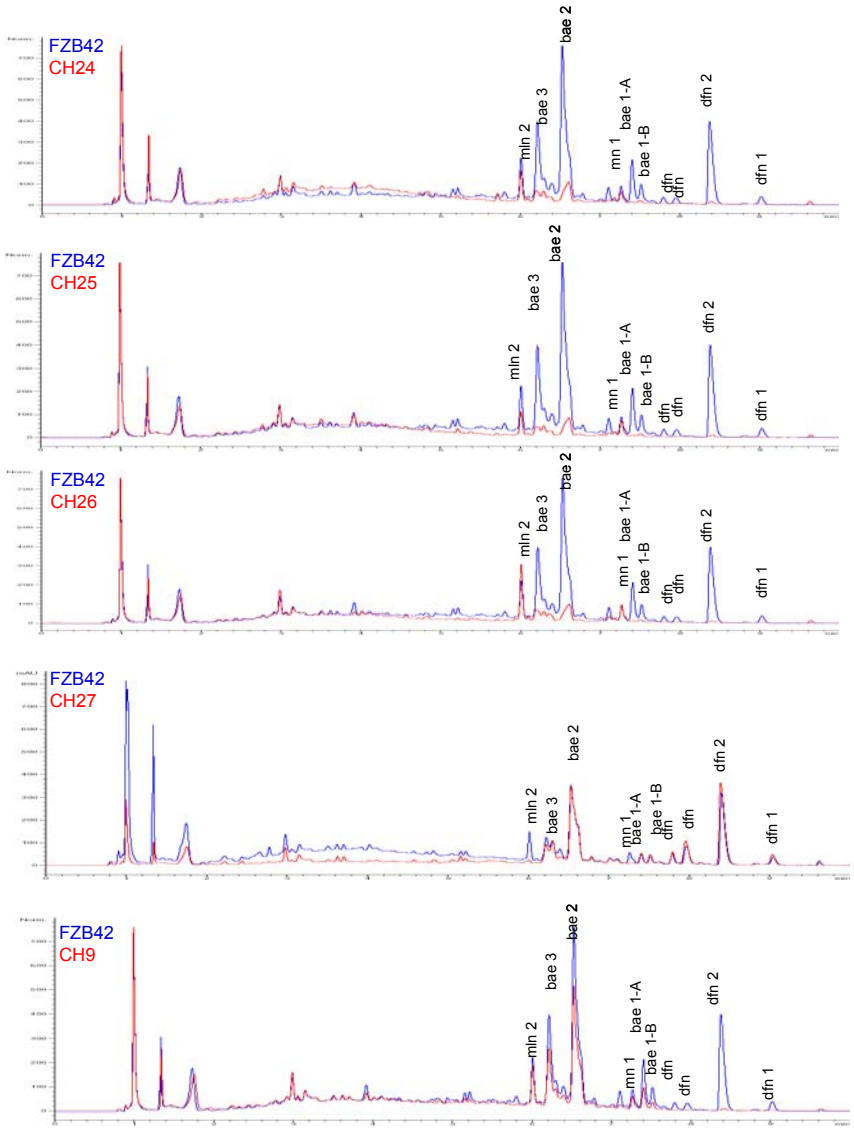


Supplementary figure **3** dfn biosynthesis pathway [**Lit. 7**]:



* Enzym e functions necessary but not detectable inside the corresponding m odules are indicated by red filled circles. The thiola tion do mains (T, ACP) predic ted to be ina ctive are not filled in the box. The KS domains without elongation function are not filled in the box. The red circle shows the postion where the cyclization of dfn happens. The green circel shows the –OH group where the phosphate group combines.

Supplementary figure **4**. Comparison of the polyketides production in the mutants with the different AT-domain deletion.



Supplementary table 1. Drug efflux protein in the genome of Bacillus amyloliquefaciens FZB42:

Gene 5'-end (bp)	3'-end (bp)	family TC No.	Product function		Note
Mdr 3053 85	306917	2.a.1.2.8	Multidrug-ef flux transporter	multidrug efflux	
Blt 1053 467	1054663	2.A.1.2.8	Multidrug resistance protein II	multidrug efflux	
ydgK 5787 43	579952	2.A.1.2.X	Putative bicyclomycin resistance protein	multidrug efflux	
yybF1 2544 377	2545583	2.A.1.2.X	Hypothetical transport protein	drug efflux	
RBAM03031 3074 660	3075856	2.A.1.2.X	Multidrug resistance protein	drug efflux	
RBAM03600 3599 461	3600661	2.A.1.2.X	Putative drug resistance transporter	multidrug efflux	
yybF 3876 926	3878128	2.A.1.2.X	Conserved hypothetical protein	drug efflux	
lmrB 269539	270972	2.A.1.3.X	Drug-export protein	drug efflux	
ycnB 4084 48	409872	2.A.1.3.X	Putative multidrug resistance protein	multidrug efflux	
yfnC 7450 17	746246	2.A.1.3.X	Putative fosfidomycin resistance protein	multidrug efflux	
yhcA 9092 94	910887	2.A.1.3.X	Hypothetical transport protein	multidrug efflux	
RBAM02833 2884 258	2885923	2.A.1.3.X	Putative permease MDR type	drug efflux	
ydhL 5938 06	594966	2.A.1.X.X	Conserved hypothetical protein	chloramphenicol efflux	
yfmO 7551 89	756371	2.A.1.X.X	Putative multidrug-efflux transporter	multidrug efflux	
yfiU 8471 08	848700	2.A.1.X.X	Putative multidrug-efflux transporter	multidrug efflux	
yfhI 8563 82	857579	2.A.1.X.X	Conserved hypothetical protein	metabolit efflux	
yitZ 1096 434	1096974	2.A.1.X.X	Putative multidrug resistance protein	multidrug efflux	
yubD 1240 251	1241808	2.A.1.X.X	Conserved hypothetical protein	multidrug efflux	
RBAM01682 1671 661	1672987	2.A.1.X.X	Conserved hypothetical protein	unknown	
ytbD 2732 693	2733890	2.A.1.X.X	Putative arabinose efflux permease	multidrg efflux	
yttB 2855 838	2857022	2.A.1.X.X	Conserved hypothetical pr otein protein	multidrg efflux	

Gene 5'-end (bp)	3'-end (bp)	family TC No.	Product function		Note
yuxJ 2974	106	2975243	2.A.1.X.X	Conserved hypothetical protein	multidrg efflux	
yusP 31	11720	3113313	2.A.1.X.X	Putative multidrug-efflux transporter	multidrg efflux	
yvkA 3350	957	3352292	2.A.1.X.X	Conserved hypothetical protein	multidrug efflux	
ywoG	3478466	3479660	2.A.1.X.X	Putative transporter	drug efflux	
ywoD 3481	317	3482681	2.A.1.X.X	Conserved hypothetical protein	drug efflux	
ywfA 3598	000	3599238	2.A.1.X.X	Conserved hypothetical protein	drug efflux	
YdfJ 5707	40	572891	2.A.6.X.X	Putative antibiotic transport-associated protein	unknown	
yojI 2043	242	2044598	2.A.66.1.X	Putative multidrug resistance protein	drug efflux	
ykkC 1254	881	1255217	2.A.7.1.5	Putative Multidrug resistance protein	multidrug efflux	ykkC/ykkD (SMR)*
ykkD 1255	219	1255531	2.A.7.1.5	Putative Multidrug resistance protein	multidrug efflux	ykkC/ykkD (SMR)*
ebrB 1779	353	1779706	2.A.7.1.6	Multidrug resistance protein	multidrug efflux	ebrA/ebrB (SMR)*
ebrA 1779	724	1780038	2.A.7.1.6	Multidrug resistance protein	multidrug efflux	ebrA/ebrB (SMR)*
yvaD 4688	08	469215	2.A.7.1.X	Hypothetical protein	multidrug efflux	yvaE/yvaD (SMR)*
yvaE 4692	12	469577	2.A.7.1.X	Putative multidrug-efflux transporter	multidrug efflux	yvaE/yvaD (SMR)*
yvdR 5980	85	598399	2.A.7.1.X	Conserved hypothetical protein	metabolite-drug efflux	yvdS/yvdR (SMR)*
yvdS 5983	99	598728	2.A.7.1.X	Conserved hypothetical protein	metabolite-drug efflux	yvdS/yvdR (SMR)*
ycxC 3719	86	372948	2.A.7.3.X	Conserved hypothetical protein	metabolite-drug efflux	ycxC/yyaM
ydzE 5847	34	585604	2.A.7.3.X	Hypothetical protein	metabolite-drug efflux	ybfH/ydzE
yojE 2046	866	2047762	2.A.7.3.X	Hypothetical protein	metabolite-drug efflux	yojE
yvbV 3249	444	3250359	2.A.7.3.X	Conserved hypothetical protein	metabolite-drug efflux	yvbV/yoav
yxxF 3734	444	3735355	2.A.7.3.X	Conserved hypothetical protein	metabolite-drug efflux	ydeD/yxxF
RBAM01170 1	158999	1159721	3.A.1	Putative ABC- type multidrug transport system, permease	unique RBAM	01171(A)/RBAM01170(M)
bceB 2858	044	2859984	3.A.1	Bacitracin export permease protein	Bacitracin	bceA(A)/bceB (M)

Gene 5'-end (bp)	3'-end (bp)	family TC No.	Product function	Note
bceA 2859 981	2860733	3.A.1	Bacitracin export ATP-binding protein	Bacitracin bceA(A)/bceB (M)
RBAM03133 3165 194	3166127	3.A.1	Putative multidrug ABC transporter , ATPase component	unique RBAM 03133(A)/RBAM03134(M) /RBAM03135(M)
RBAM03134 3166 153	3167245	3.A.1	Putative multidrug ABC transporter , permease	unique RBAM 03133(A)/RBAM03134(M) /RBAM03135(M)
RBAM03135 3167 249	3168392	3.A.1	Putative multidrug ABC transporter , permease	unique RBAM 03133(A)/RBAM03134(M) /RBAM03135(M)
spaG 3210 566	3211336	3.A.1	Lantibiotic ABC transporter permease	subtilin spaF(A)/spaE (M)/spaG(M)
spaE 321 1333	3212067	3.A.1	Lantibiotic ABC transporter permease	subtilin spaF(A)/spaE (M)/spaG(M)
spaF 3212 064	3212777	3.A.1	Lantibiotic transporter (A TP-binding protein)	subtilin spaF(A)/spaE (M)/spaG(M)
mrsF 3770 199	3771108	3.A.1	Putative ABC transporter, A TP-binding protein	mersacidin m rsF(A)/mrsG(M)/mrsE(M)
mrsG 3771 107	3771863	3.A.1	Putative ABC-transporter inte gral membrane protein	mersacidin m rsF(A)/mrsG(M)/mrsE(M)
mrsE 3771 882	3772614	3.A.1	Putative ABC-transporter inte gral membrane protein	mersacidin m rsF(A)/mrsG(M)/mrsE(M)
yfiN1 2259 17	227057	3.A.1.X.X	Putative ABC-type multidrug transport system, per mease component	unknown y fiL(A)/yfiM(M)/yfiN1(M)

A= ATP binding domain
M= Transmembrane domain

6 Appendix 2:

My publications directly linked with this PhD thesis:

1. **Chen XH**, Koumoutsi A, Scholz R, Borriss R. Genome analysis of Bacillus amyloliquefaciens FZB42 reveals its potential for biocontrol of plant pathogens. *accepted* J Biotechnol. 2009.

2. **Chen XH**, Scholz R, Borriss M, Junge H, Mögel G, Kunz S, Borriss R. Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. *accepted* J. Biotechnol. 2009.

3. Moldenhauer J, **Chen XH**, Borriss R, Piel J. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angew Chem Int Ed Engl. 2007; 46(43):8195-7.

4. Schneider K*, **Chen XH***, Vater J, Franke P, Nicholson G, Borriss R, Süssmuth RD. Macrolactin is the polyketide biosynthesis product of the pks2 cluster of Bacillus amyloliquefaciens FZB42. J Nat Prod. 2007 Sep; 70(9):1417-23. Epub 2007 Sep 11. * These authors contributed equally.

5. Koumoutsi A, **Chen XH**, Vater J, Borriss R. DegU and YczE positively regulate the synthesis of bacillomycin D by Bacillus amyloliquefaciens strain FZB42. Appl Environ Microbiol. 2007 Nov;73(21):6953-64. Epub 2007 Sep 7.

6. **Chen XH**, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess WR, Reva O, Junge H, Voigt B, Jungblut PR, Vater J, Süssmuth R, Liesegang H, Strittmatter A, Gottschalk G, Borriss R. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium Bacillus amyloliquefaciens FZB42. Nat Biotechnol. 2007 Sep; 25(9):1007-14. Epub 2007 Aug 19.)

7. **Chen XH**, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süssmuth RD, Borriss R. Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. J Bacteriol. 2006 Jun; 188(11):4024-36.

8. Koumoutsi A, **Chen XH**, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in Bacillus amyloliquefaciens strain FZB42. J Bacteriol. 2004 Feb; 186(4):1084-96.

9. Patent title: Sequences for the synthesis of antibacterial polyketides, inventors: Prof. Dr. Rainer Borriss, **Xiao-Hua Chen**, Alexandra Koumoutsi, Dr. Joachim Vater, international publication date: 08.12.2005, international publication number: WO 2005/116061 A2.

Selbständigkeitserklärung

Hiermit versichere ich, die vorliegende Dissert ation selbstständig verf asst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben.

02.12.2008

Xiaohua Chen